CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

2·1

ERNST PULGRAM

Introduction to the Spectrography of Speech

MOUTON & CO - 'S-GRAVENHAGE

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

INTRODUCTION TO THE SPECTROGRAPHY OF SPEECH

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

JANUA LINGUARUM

STUDIA MEMORIAE NICOLAI VAN WIJK DEDICATA

edenda curat

CORNELIS H. VAN SCHOONEVELD
LEIDEN

NR. VII

1959
MOUTON & CO · 'S-GRAVENHAGE

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

INTRODUCTION TO THE SPECTROGRAPHY OF SPEECH

by

ERNST PULGRAM
UNIVERSITY OF MICHIGAN

1959
MOUTON & CO · 'S-GRAVENHAGE

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

© 1959 Mouton & Co., Publishers, The Hague, The Netherlands.

No part of this book may be translated or reproduced, by print, photoprint, microfilm, or any other means without writtten permission from the publishers.

Vidya Niwas Mishra
Sanskrit Department
University of Gorakhpur,
GORAKHPUR.

डा० विद्या निवास मिश्र
प्राध्यापक, भाषा विज्ञान
निजी पुस्तकालय
पुस्तक संख्या.....

Printed in The Netherlands by
Mouton & Co., Printers, The Hague

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

PREFACE

The purpose of this book is to inform linguists on a subject of steadily increasing concern to them, spectrography. It is meant to fill a gap in the literature of linguistics, in that it introduces the linguist who is unable to follow the technical discussions and to peruse the specialized literature of physics and acoustics, to the principles involved in the nature and the functioning of the spectrograph. But I hope this volume may also be useful to some acousticians and communication engineers interested in language, in that it presents to them those tenets and theories of linguistics which are necessary for bringing about an intelligent and proper coordination and correlation of acoustics, spectrography, and linguistics. I have therefore taken care at all points of my text to be linguist and spectrographer and acoustician at the same time, and I have insisted on never dealing with acoustics without continued reference to linguistics, and vice versa.

I am a linguist, however, and not an acoustician. Hence my discussion of acoustic problems as such consists of what I have learned, and not of original research. But I have thought it useful to compile whatever I have learned, and to make it available in cohesive form to other linguists who might be as uninformed on this subject and as eager to know more about it as I was several years ago. It is my hope and my aim that by using the present book they may be spared the laborious and time-consuming procedure of false starts, wrong paths, and incomplete results with which I struggled.

I acknowledge with pleasure and gratitude the help I received from two experts above all, Martin Joos and Gordon E. Peterson. It was Joos who gave us the first technical and

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

linguistically useful book on *Acoustic Phonetics* (Language Monograph No. 23, 1948), from which I derived much general information, and specifically, with the permission of Joos and the editor of *Language*, some of the illustrations and charts used in the following pages (Figs. 16, 17, 18 are redrawn after Joos; Table II is copied from Joos).

It has been my particular good fortune to be, at the University of Michigan, a colleague of Gordon E. Peterson, who has with unfailing patience and kindness instructed me and encouraged me and many a time retrieved me from blind alleys—to say nothing of his answering certain questions of mine which must have seemed, to put it mildly, naive. He also put at my disposal the resources of his outstanding laboratory, and several illustrations (Figs. 19, 28, 31, 32) are his work. Without his proximity and continued assistance I could not have finished my manuscript.

Mr William S. Y. Wang, of the staff of Peterson's laboratory, has given me many hours of his time for discussion, and he has advised me and corrected me on many points, for which I express my sincere thanks. He also made several spectrograms for me (Figs. 21, 29, 30). I am also grateful to Dr Ilse Lehiste, of the same laboratory, who was good enough to read my manuscript and who made a number of suggestions and corrections.

Fig. 20 is a reproduction of a drawing made by the Bell Telephone Laboratories.

Needless to say, none of the persons just named is in any manner responsible for any errors or inaccuracies that remain in this book. And for all I know, owing to the rapidity of development in this field, some details may become obsolete between the conclusion of the manuscript and the publication of the book. I hope, however, that this will not affect the principal and general points I wish to make.

The extent of my indebtedness to other scholars is measured by the Bibliography. The expert reader will easily identify my chief sources, although I have refrained from burdening this

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

text with footnotes; and the non-expert will, I trust, turn for further readings to the sources I myself used and listed, and thus discover for himself to whom I owe what.

Over the course of several years I have successfully and profitably incorporated increasing quantities of spectrography in my phonetics lectures, and I found, not to my surprise, that conscientious teaching is an excellent way of learning. I also had the pleasure of giving an entire course of lectures at the University of Florence, Italy, in 1956-57 on *Spettrografia del linguaggio,* and I found, again not to my surprise, that switching from one linguistic code to another is conducive to self-critique, fresh views, and greater lucidity. I hope that the potential benefits of these teaching experiences are not wholly lost in my text.

Although the chief purpose of this book is to state objective scientific facts, I have not been able, nor particularly anxious, to conceal the subjective fact that I have opinions. I am rather an adherent of what has been called "God's truth" linguistics, and not of "hocus pocus" linguistics, in that I believe (1) that linguistics is not merely, or primarily, a series of operations which orders a mass of data into a structure, but a science which examines the available data in order to discover the underlying structure; (2) that phonetic data are not irrelevant in the discovery of the phonemic structure of a language, and that they cannot be dispensed with in favor of purely distributional criteria; (3) that phonemic contrasts are statable in terms of binary oppositions; (4) that meaning is a linguistic and not a metalinguistic property of language.

These opinions I now hold as a consequence, at least in part, of my studies in spectrography. To the extent that I wish to convert others to similar views or guide beginners to this end, this book is not innocent of tendentiousness.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

TABLE OF CONTENTS

Preface 5

ACOUSTICS

I. Sounds, speech sounds 11
II. Tones 24
III. Oscillograms 35
IV. Glottal tone 43
V. Phase 46
VI. Spectrum 50
VII. Resonance, reinforcement 55
VIII. Damping 60
IX. Filtering 65

PHONETICS, PHONEMICS

X. Articulation 69
XI. Phonemization 78
XII. Vowel phonation 85

SPECTROPHONETICS

XIII. Spectrophones 89
XIV. Spectrograph 97
XV. 18~ filter 106
XVI. 50~, 300~ filter 119
XVII. Formants 126

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

SPECTROPHONEMICS

XVIII. Resonance bands 133
XIX. Spectrophonemes (vowels) 145
XX. Spectrophonemes (non-vowels) 157
XXI. Spectrographemes 165

Selected bibliography 171

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

ACOUSTICS

I

SOUNDS, SPEECH SOUNDS

When one plucks a violin string or draws a resined bow across it, when one strikes a tuning fork or a drum, when one blows into a trumpet or over the rim of a bottle, one produces, and hears, a sound. This auditory impression is due to vibrations of the string, the surface, or the column of air, a motion whose impulses are communicated to the surrounding medium, which is normally air.

The vibrations which are conducted through the air are not equivalent to a movement of the air: sound waves are not wind, and wind is not necessary for their propagation. The vibrations of the sound-producing body merely exert an impulse upon the contiguous particles of the sound-carrying atmosphere. Each particle thus propelled crowds upon its neighbors, creating a compression of air before it and a rarefaction behind it. The neighboring particles are pushed on and cause the compression to progress in the direction in which the sound is traveling. The particles are presently sucked back into the rarefaction which their movement has created behind them, indeed they swing somewhat beyond their original position because of inertia. If in the meantime the sound-producing body has completed another cycle of oscillation, they are once more impelled on their way and the whole performance is repeated. The process continues until the original source of power, the sound-producing body, its energy spent or cut off, stops moving; whereupon after some further pendulum motions and recoils, a relative equilibrium is restored.

Wherever the air vibrations impinge upon an obstacle some of their energy is expended. If the obstacle is rigid and impervious to the force by which it is struck, it may simply reflect

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the oscillations of the air and be little or not at all affected by them; but if the obstacle is responsive and easily set in motion, then it will in turn commence to vibrate under the impact. The membranes of telephones and microphones are obstacles of this type receptive to air waves. The energy that strikes them is translated, or encoded, into electric currents, which are in the receiver electronically decoded again into sound waves. These are transmitted to the ear, where they impinge, as do of course other vibrations coming directly from their sources without intermediate electronic transmission, upon a membrane which is part of the human hearing apparatus, the eardrum. Through it, and a very sensitive and complex set of conducting devices located behind it in the middle and inner ear, the oscillations are this time transformed into nervous impulses, which in turn are conveyed to the brain and perceived by the hearer as sound. Acoustically speaking, then, the impression of sound is caused by the vibrations of a sound-producing body.

The perception of sound endures until the last wave of compression and rarefaction has impinged upon the eardrum, at which point relative silence (scarcely ever, in our world, complete silence) is restored. No particle set in motion by the movement of the sound-producing body actually travels the whole distance from its original position to the eardrum. In other words, the transmitting medium does not move along with what one calls the sound wave. This phenomenon can be compared with the progress of waves in water. (But air waves are not exactly like water waves, the first being longitudinal, the second, like light waves, transversal. In a longitudinal wave the amplitude is parallel to the motion of the wave, in a transversal wave it is at right angles.) Although waves in liquids give the impression of traveling in some direction, there really is no current necessarily combined with their movement. A piece of cork placed in the midst of concentrically spreading waves in a quiet pool of water will merely bob up and down (actually it has a vertical circular motion) but it will not travel appreciably in the direction of the waves. (An imaginary cork

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

floating in air, however, would, if set in motion by a sound wave, which is longitudinal, not bob up and down but move back and forth in the direction of the spreading sound.) Currents and waves operate essentially independently of one another, as do wind and sound.

A vibrating body that produces a sound need not be a musical instrument which brings forth a musical tone of measurable pitch, but it can be any object which may be caused to vibrate if struck or scratched or otherwise set wholly or partly in motion. Some of the percussion instruments of an orchestra fit into this category, where they are used chiefly to beat out a rhythm rather than to add to the melody. Such sounds are called noises, which differ from tones by not having a measurable pitch on the musical scale; they are therefore musically indifferent or sometimes even unpleasant.

But not all vibrations in nature are actually perceived by the human ear and brain. Their physical properties may cause some of them to lie beyond the threshold of human perception in terms of either loudness or pitch. A single waving blade of grass undoubtedly causes air vibrations, and consequently, by definition, a sound. But this sound is not loud enough, nor possibly of suitable frequency, to be perceived by the naked ear. Nor can man 'hear the grass grow', though this movement, too, inevitably produces some noise. The auditory organs of non-humans may have different thresholds. Some animals are known to have a better sense of hearing than man. As for the perception of pitch, a dog can hear a dogwhistle blown by his master, which remains inaudible to humans because of its high pitch. (One could argue that what cannot be heard is not a sound. But given the varying thresholds among different individuals and different ages, to say nothing of those of non-humans, a sound is better defined as vibration rather than in terms of perceptibility.)

The human vocal apparatus (*Fig. 1*) is a complex assembly for producing sounds, both tones and noises. I shall briefly describe its parts and their function.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

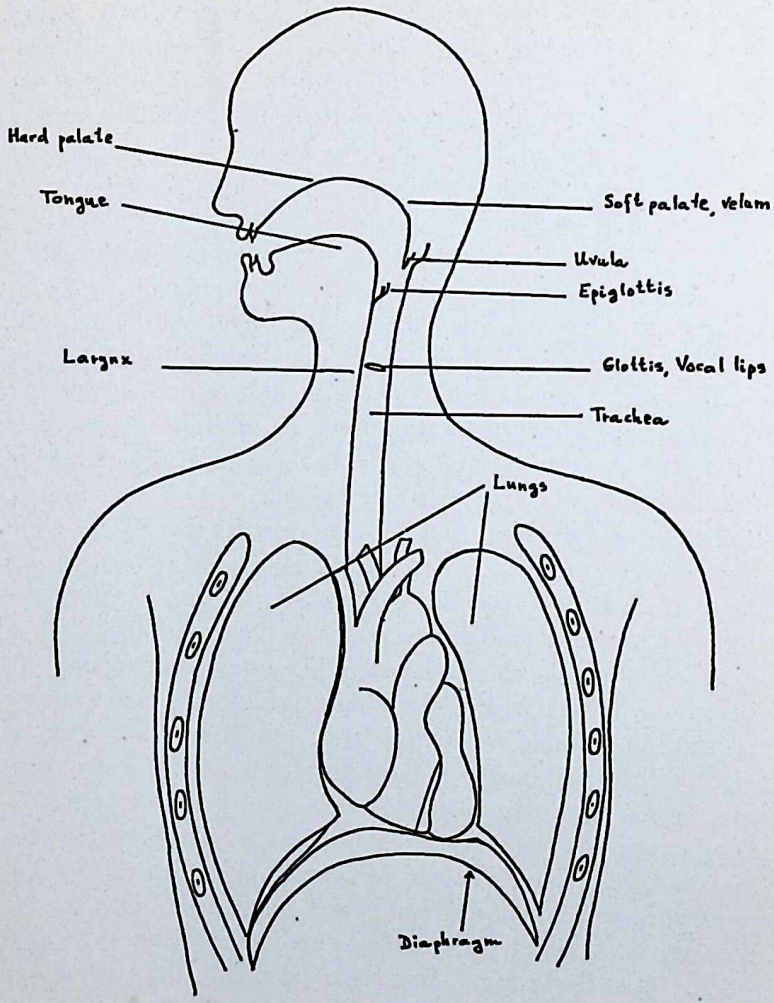


Fig. 1. The human vocal apparatus.

The diaphragm is a muscular membrane (present only in humans and other mammals) which separates the thorax from the abdomen and which by its rhythmic up-and-down movements assists in causing the lungs to contract and expand, that

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

is, to exhale and inhale the stream of air needed for breathing. During expiration (and rarely during inspiration) the moving air provides the energy which sets into vibration some of the movable parts of the speech apparatus, or the speech organs.

Of great importance among these organs are the vocal cords, which I prefer to term vocal lips since they are not cords or strings suspended like the strings of a violin, but rather form a mouth-like orifice, the so-called glottis, above the trachea (windpipe) at the larynx (voice box). The glottis may be opened or closed at will, and the vocal lips, like the lips of the mouth, may at will be relaxed or tensed, opened or closed, so that they are either set to vibrating by the escaping air or allowed to remain lax and at rest without obstructing the passing air. When the vocal lips vibrate they contribute significantly to the articulation of vowels and voiced ('soft') consonants (*b, d, g, v, z, m, n,* etc.); they do not vibrate when an unvoiced, or voiceless, consonant (*p, t, k, f, s,* etc.) is articulated.

Some languages (Japanese, for example) have voiceless or whispered vowels. According to the definition of vowels just given this implies a contradiction in the terminology. However, these sounds may properly be called vowels, at least in terms of phonemics, since they fulfill the same structural function as do the voiced vowels. We are faced here with one of those unfortunate situations in which the terminology of phonetics and that of phonemics are at odds: the classification of a phenomenon is performed, in the first, in terms of physical properties (articulatory, acoustic, etc.), in the second, in terms of its function in the system.

Also whispered speech is performed without activation of the vocal lips. The recognition by the hearer of what would normally, in English (and many other languages), be significant oppositions between voiced and voiceless speech sounds, such as in the pairs *dip* and *tip, cab* and *gap,* is then accomplished by sets of accompanying distinctive features (lax *vs.* tense, lenis *vs.* fortis, etc.). In unusual circumstances, then, the conventional principal distinction may be omitted and secondary

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

features are made to serve in its place. (This implies a certain redundancy of signaling, of which more will be said.) But these are admittedly deviations from the norm, which the hearer recognizes as such and for which he makes allowances and automatic adjustments.

From this type of whisper the so-called stage whisper must be distinguished. The latter is not really a whisper at all but merely normal speech of reduced loudness and increased breathiness. That is to say, the volume of air which is employed in the articulation of stage whisper is much larger than normal, which imparts to the sounds a false whispered quality, a breeziness reminiscent of true whispering.

The tongue, the teeth, the lips, the palate, and the lower jaw contribute to the modification of the tone produced by the vibrations of the vocal lips, or, in the absence of this tone, cause themselves the noises which we classify and identify as unvoiced speech sounds. Since most of the vocal organs (especially the tongue, which is among the most mobile and agile and sensitive muscular organs of the body) can be moved in many ways by volitional muscular activity, a human is capable of creating many different sounds. He can change at will the relative size and shape of the supraglottal cavities (oral, or oral and nasal—but note that the nasal cavity by itself is not capable of being modified in size and shape since, unlike the mouth, it contains no volitionally movable organs) by altering the relative positions of the speech organs; he can at will activate most of these organs, at least in relation to one another, or let them rest, singly or in groups. Since it takes but minute alterations in the positions of the vocal organs to produce distinguishable varities of sounds (the reasons for this shall become clear later), the human vocal apparatus is a very delicate instrument capable of originating an enormous variety of distinct tones and noises.

Of this vast potential of articulations each language employs only a rather restricted number of classes of articulations, ranging approximately between two and six dozen. Every such

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

class is called a PHONEME. Different languages employ different numbers and types of phonemes, occurring in different distribution, that is to say, following different rules of permissible sequences and positions in relation to pauses. For example, English has a /θ/, spelled *th*, German has not; German has a /χ/, spelled *ch*, English has not; Spanish has only five vowels, English has nine (or, some say, ten or eleven); a German utterance may begin (post-pausal position, word-initially) with /pfl/ as in *Pflaume* 'plum', an English utterance may not; a German utterance may end (pre-pausal position, word-finally) with /rbst/ as in *Herbst* 'autumn', an Italian utterance may not; English has in medial and final but not in initial position a /ŋ/, as in *singing*, French does not have it at all; French has in medial and final but not in initial position a /ɲ/ as in *soigner* 'to cure' and *campagne* 'country', English has no such phoneme (although the sound occurs as a variant, called allophone, of /n/, as in *onion*).

Every speaker acquires the stock of phonemes of his native language together with the rules of their distribution from infancy, by imitation and trial and error. The child develops in this domain of muscular activity a kinaesthetic sense, a self-regulating feedback, so that he learns to produce automatically sounds belonging severally to the proper phonemes. This becomes so ingrained a part of man's linguistic cultural behavior that the layman is often tempted to think of it as pertaining to his biological inheritance. (The term 'native language' reinforces the notion of language as being something born with the individual.) Such a view may even lead to the belief that a native speaker of English who descends from, say, French forebears inherits a greater talent for learning French than for Italian or Spanish, or that he learns French more quickly and easily than does a person of different ancestry. There may exist in such a student of language not a greater aptitude but a predilection for French; if so, it has its roots in cultural preferences or traditions or family sentiments, not in biological inheritance. A language, and its phonemes, and other struc-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

tural parts, are no more biologically inherited than a person's habits of dressing or his table manners. They are, instead, cultural features acquired from the environment, and they vary therefore, like fashions of eating and dressing, from society to society, from culture to culture, from period to period. But sameness of culture need not coincide with sameness of language, nor variety with variety.

The faculty of speech, however, irrespective of language, is indeed 'native', being an innate part of the real heritage of man. Man is the only speaking animal, he alone has the requisite cerebral and nervous constitution allowing him this peculiarly human behavior. He has obtained this equipment by way of evolutionary mutations, we do not know why or how. Animals may communicate with one another, but they do not talk, and they do not acquire, store, and transmit accumulated knowledge by means of language. (And they cannot make negative statements—"Something is not"—, or lie.) Man's special linguistic aptitude may have something to do with his ability to use extra-corporeal tools, but the origins of speech and the conditions of human proto-language remain an unsolved mystery.

It is worth noting that man did not develop a new set of physical attributes to permit him the production of speech. All parts of his vocal apparatus are merely adapted to that special task, while they are fundamentally designed to serve primary vital, biological, physiological functions. The diaphragm, the lungs, the trachea, the pharynx obviously are needed for breathing, for providing the organism with oxygen that nourishes the blood-stream. The vocal lips, the cartilages of the larynx protecting them (the largest of which is the Adam's apple), and the epiglottis (a shoehorn-shaped cartilage projecting over the trachea at the base of the tongue) protect the bronchi and the lungs from intrusion of foreign matter. The mouth, with the tongue, teeth, jaws, lips, serves to grasp, hold, and masticate food. The tongue is also the seat of the sense of taste. The nose's primary function is to filter, warm, and moisten the air before it enters the lungs, and to convey scents to the olfactory centers. And

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Table I

PHONEMES OF AMERICAN ENGLISH

	PLOSIVE		CONTINUANT				
			FRICATIVE		Nasal	Liquid	Spirant
	voiced	voiceless	voiced	voiceless	voiced	voiced	voiceless
bilabial	b	p	w		m		
labiodental			v	f			
dental	d	t	z	s	n	l, r	
alveolar			ʒ	ʃ			
interdental			ð	θ			
palatal, velar	g	k	j		ŋ		
pharyngeal							h

Examples: *b*ane *p*ane *w*ane *m*ane
*v*ane *f*ame
*D*ane *t*ame *z*any *s*ame *n*ame *l*ane, *r*ain
a*z*ure *sh*ame
*th*an *th*in
*g*ain *c*ane *y*am ba*ng*
*h*ang

---

	front		center		back	high	closed
	i				u		
unrounded	I				U		rounded
	e					medium	
			ʌ		o		
		ɛ					
		æ		ɔ			
			a			low	open

Examples: b*ea*t b*oo*t
b*i*t
b*a*te b*u*d B*u*ddha
b*e*t b*oa*t
b*a*t b*ough*t
f*a*ther

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

also the ear, by means of which we do not speak but only hear speech, is associated as well with the sense of equilibrium.

Persons using the same set of phonemes occurring in the same distribution, together with other sets of sames on different levels (forms, syntax, lexical units), are said to speak the same dialect or language, or to belong to the same speech community.

Table I lists the phonemes of American English according to their manner and place of articulation. Not all linguists are agreed on the phonemic structure of American English—*if* there is only *one* American English. But it is not germane to my purpose to deal with this controversy here. I shall only note, without enlarging upon the current controversy among two schools of linguists, whom one might name physicalists and relationalists (distributionalists), that I think physical considerations of the essence in linguistic analysis, especially on the phonemic level, since language implies physical data, which can be observed and measured according to their physical structure. Having obtained the physical data, the linguist must classify them, and in this step he inevitably must take account of the fact that the data are also cultural, in addition to being acoustic, phenomena. But totally excluding the physical side seems, to say the least, awkward.

Throughout this book, then, I shall take the position, which I have at least in part adopted in consequence of learning to view language spectrographically, that neither pure physicalism nor intransigent relationalism can provide all the answers we need to know about language, but that, rather, the physical and the cultural aspects of language must be treated jointly.

One of the basic tenets of phonemic theory states, for example, that the initial [l] and the final [ɫ], two phonetically different sounds, of the English word *lull* are to be classified as positional variants, as allophones, of one phoneme /l/ and not as two separate phonemes, because one occurs predictably in one, the other in the other position, but neither in both positions. Hence the statement /l/, if qualified for phonetic purposes with reference to position, is unambiguous and com-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

plete. We say that sounds occurring in complementary distribution are allophones of a phoneme, and not separate phonemes. In addition—and this is, to my mind, of greatest weight—a native informant, if asked whether the word *lull* has or has not the same sound at the beginning and at the end, will invariably affirm that it has. And finally, [l] and [ɫ] are phonetically similar.

To extend, however, the criterion of complementary distribution to [h] and [ɫ], and to declare them members of the same phoneme on the grounds that [h] never occurs finally and [ɫ] always, does violence to both the reaction of the native speaker (who will never admit that *hell* begins and ends with the same sound) and the requirement of phonetic similarity.

True, similarity is relative: how distant may two sounds be in terms of phonetics and still be called similar? The case of [h] and [ɫ] is easily solved since the [ɫ] belongs, on phonetic evidence, together with other [l]-like allophones to the phoneme /l/, regardless of its distributional relation to [h]. And I know of no phonemicist, no matter how distributionally inclined, who classifies differently—implicitly admitting thereby his consideration of phonetic similarity. I do not know what such an analyst would do if [ɫ] were the only allophone of /l/; he might then possibly, on distributional grounds, put [l] and [h] actually into one phoneme. I myself should not do this, for reasons already stated. Furthermore, spectrographic analysis will furnish, as we shall see, additional reasons for using both distributional and phonetic criteria in phonemic analysis and classification.

Of course, if phonemics is a bookkeeping system concerned frankly and exclusively with matters distributional, then there is nothing to prevent anyone from classifying [h] and [ɫ] as belonging to one phoneme. But since I prefer, as I said in the Preface, to apply phonemic analysis as an operation which examines the available data in order to discover the underlying structure, and not as an operation which forces a mass of data into a structure, the physical content of each item is of

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the essence. The decisive point here is not, then, to decide whether [h] and [ɦ] 'really' are, or are not, phonemic sames, but whether I do or do not define phoneme and phonemics with the inclusion of criteria pertaining to the informant's reaction and phonetic similarity. I, and many others, prefer the latter, and my subsequent spectrographic discussion will attempt to justify further, or at least to state further reasons for, this preference.

One must not forget that, physically, speech sounds are tones and noises like other tones and noises that occur in nature. They are distinguished from the totality of the sounds that surround us by two factors. First, they are produced by the human vocal apparatus (although there are ways of making synthetic speech sounds: but they are imitations of the natural ones, the closer the better). Second, they are sounds selected from among the mass, and specifically elevated to their peculiar linguistic position and function. This selection is due to a cultural agreement, to an unwritten contract among the members of a speech community to use a certain number and certain types of possible sound classes as the phonemes of their language. All members of a speech community are thus in possession of a manageably small number of generally valid rules of acoustic behavior in regard to language. According to these precepts, which they acquire while growing up by imitating models they hear, children learn this peculiar game of making noises with their faces. All members of the community, provided they are normally endowed in body and mind, attain such efficiency and proficiency in this highly complex activity that speech becomes an intelligent and successful instrument of communication and cooperation.

Since speech sounds are, then, basically not different from other sounds, there is every reason for examining them in the same way as other sounds that one hears, at least as regards their physical, acoustic properties. But being at the same time intrinsic parts of an elaborate structure of cultural behavior patterns, of a language, speech sounds may and must be

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

examined also from the point of view of the cultural, behavioral sciences. Since I am a linguist and not an acoustician, and since my and, I suppose, my readers' interest is linguistic and not primarily acoustic, linguistic considerations will be paramount in the acoustic analysis and visualization of speech described in the following pages. Consequently I must aim, in the description and construction of the analyzing machine, the spectrograph, for performance and results which are above all linguistically meaningful and relevant (although this will by no means exhaust the potentialities of the spectrograph for acoustic research).

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

II

TONES

If I strike a tuning fork, the ends of the two prongs vibrate as shown in *Fig. 2.* The arrows indicate the maximum width

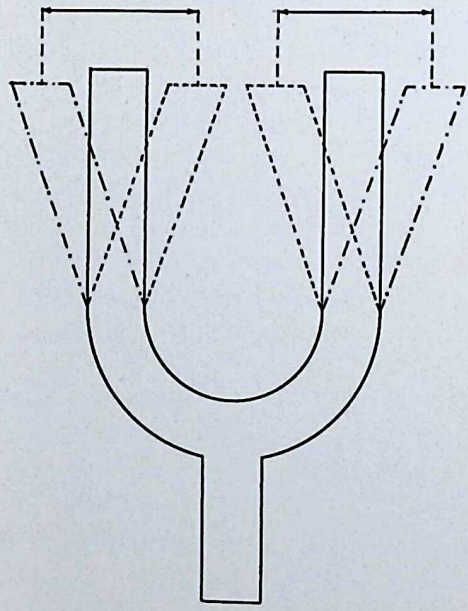


Fig. 2. Vibrating tuning fork.

of vibration, which is proportionate to the force with which the fork is struck and which depends also, of course, on the fork's physical properties—its material, weight, and shape. The width of the oscillation is called AMPLITUDE. The tips of the fork swing at a certain speed, which can be measured per unit of time. One may find, for example, that the tips of the fork

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

move from a position of rest to one extreme position, then the other, and back again to the position of rest two-hundred times per second. This measurement is called FREQUENCY and is expressed in cycles per second, thus: 200 cps, or 200∼. The number of cps depends on the length of the prongs of the tuning fork (and, again, its physical properties): the longer the prongs, the more time is required for the tips to traverse a cycle, that is, the lower the frequency of the tuning fork. The EXTENSITY or DURATION of the movement is measured in time: the fork vibrates for a certain number of seconds.

Since the vibrations are taken up by the air and spread through a series of compressions and rarefactions of density which originate at the tuning fork and cause the pendulum movement of air particles, and since upon striking the eardrum they are eventually perceived as sounds, the question may be asked in what manner each of the quantities just named is heard. The amplitude, which corresponds physically to the SOUND PRESSURE, that is, the deviation in pressure from the surrounding atmospheric pressure due to the displacement of the air particles, measured in dynes, corresponds perceptually to the LOUDNESS of the sound, which is measured in sones. This quantity is one of perception, and therefore subjective. Objectively the phenomenon is termed acoustic POWER and is measured in watts, the same kind of unit in which electric power consumption is computed. The necessity for distinguishing loudness and power is due to the behavior of the ear, which does not translate an increase of power into an arithmetically equal increase in loudness. Moreover, the distortion is not the same at all levels of frequency. For example, at low power levels, high frequency tones sound louder than low frequency tones of the same power; but at high power levels, all tones of the same power sound almost equally loud, regardless of frequency; hence a change in power at low power levels produces greater change in loudness at low frequencies than at high frequencies; but a change in power at high power levels produces virtually the same change in loudness regardless of

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

frequency. Also, at the same power level, one frequency sounds at a minimum loudness, higher and lower frequencies sound at greater loudness.

The human ear is susceptible to a very wide range of acoustic powers. The loudest sound which it can perceive without pain or damage has about one million times the pressure of the feeblest audible sound. In order to handle these large figures more easily in mathematical calculations, loudness is often measured in a unit designed for this purpose which is called a decibel, abbreviated db. A db is actually a unit that measures the ratio of any two amounts of electric or acoustic energy, just as percent measures the ratio of two figures. But recent standardization has established a permanent reference scale of energy in relation to which db has acquired also an absolute meaning (as, for example, a degree on a thermometer which is calibrated according to the boiling point and freezing point of water at sea level). The standard reference value chosen for the zero point was 10^{-16} watts per square centimeter, that is, one 10-quadrillionth of a watt. This extremely small amount of energy is just above the threshold of perception of a human ear of slightly better than normal sensitivity. Also, the least difference in loudness detectable by the average hearer is approximately 1 db; but this difference seems to depend somewhat also on the type of sound under test, whether it is a tone, or a noise, and what kind of either. On this db scale, average conversation lies between 50 and 70 db. A noise louder than 120 db causes pain, continued 'ringing' in the ears, even temporary deafness. (In what follows I need to make little use of the db.)

The frequency determines the PITCH one hears. The greater the frequency, the higher the pitch. A doubling of the frequency raises a tone by one octave on the musical scale. Here, too, for the same reasons as cited in the discussion of loudness, there exists a subjective and an objective measure of quantifying pitch, the former using mels as units of measurement, the latter cycles per second.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

In speaking of amplitude—loudness and frequency—pitch we employ, then, two types of description and measurement for each: db and watt for the first, mel and cps for the second. Db and mel are measurements of perception, describing what the human observer perceives, watt and cps objective measurements of energy, describing what a non-human machine registers. The two sets are not related to one another in a simple proportion; if plotted on a system of coordinates, their relationships appear not as simple straight lines but as curves, known as psycho-acoustic curves. The anthropologist, or psychologist, or linguist operating on the cultural level (phonemicist), may be satisfied with one set; the physicist, or acoustician, or linguist operating on the natural science level (phonetician), with the other; but the experimental linguist, who, like the spectrographer of speech, is interested in the cultural *and* the physical values of the phenomena at the same time, and desires to establish a relationship between them, must interpret the evidence that comes from both sources and in both sets of terms, and must eventually describe it in a third, the psycho-acoustic or cultural-acoustic, set of terms. This is another instance pointing to the requirement of viewing language as a cultural and physical phenomenon simultaneously. Naturally, simultaneity neither implies nor invites methods or descriptions of promiscuity and confusion.

If in the place of a set of four tuning forks of various length of prongs we use four taut strings of equal length and mass, we can still produce tones of varied pitches by varying the tautness of the strings (with limits being imposed by the elasticity of the material). The tauter a string the faster it vibrates when plucked, and the higher the pitch of the resultant tone. The vocal lips can, by means of muscular action, be similarly tautened and relaxed. Since they are fleshy rather than metallic, their elasticity is high and the variety of possible pitches is great. (Note that while tautening is accompanied by a lengthening of a given pair of vocal lips, which one would normally associate with a lowering of pitch, the results of the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

increase in tautness prevail over those of the increase in length, and the tone gets higher rather than lower.)

The four nearly equally long strings of a violin are tuned in fifths, with each string separately capable of pitch adjustment by means of tautening. But in order to provide the instrument with the four basic differences of pitch without overtaxing the tensile strength of the strings, the task of tuning is not left to variegation in tautness alone but is accomplished in large degree by the use of four strings of different mass (material, thickness). Since the frequency decreases as mass increases, and vice versa, for a given length and tension, a thin wire of small mass can be more easily tautened so as to produce high tones while catgut strings (and, on bass fiddles and pianos, heavy spiral strings) of larger mass lend themselves better to the creation of low tones at a lesser degree of tension. Of course, each string can separately be caused to produce varied pitches by stopping, that is, by lengthening or shortening it through the placement of fingers along it: that is what the fiddler does with his left hand.

Each speaker possesses only one pair of vocal lips, and he has no device for stopping in his larynx. Therefore, varying the pitch of his performance in speaking or singing is always a matter of varying the tension of the vocal lips. On the other hand, different speakers have vocal lips of different length and mass, so that each speaker, like each string on an instrument, has a basic, natural range of pitches which he cannot exceed. His anatomic conditions provide then each speaker with a soprano, or alto, or baritone, or basso voice. (Women and children, generally having vocal lips of smaller mass, belong in the upper registers, men with vocal lips of greater mass belong in the lower registers.) A trained singing voice can learn to reach both higher and lower than an untrained voice generated by an anatomically equal apparatus, and to perform accurately and without wavering at a prescribed pitch. For the purpose of tautening and relaxing the vocal lips every speaker and singer has at his disposal an ingenious (if the term

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

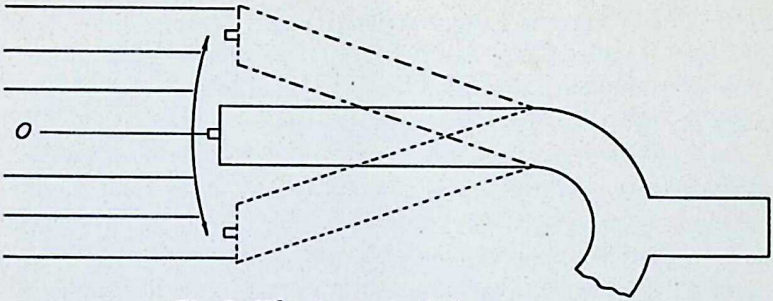


Fig. 3. Vibrating tuning fork—one prong.

is apposite here) set of muscles which he can activate at will.

Let us consider, for the sake of simplicity, the vibrations of but one prong of a tuning fork. By providing the horizontally held prong on its underside with some kind of stylus, and by pressing it (lightly, so as not to lose too much energy through friction) against a sheet of paper, we could produce a result as shown in *Fig.* 3. The tracings on the paper would of course repeat over and over the section of a circle, up and down the sheet. In order to represent this movement as a function of time, we may, instead of holding the paper quiet, pull it away at a regular speed from under the vibrating prong. Instead of running over the same course all the time, the stylus now draws a wavy line, called an oscillogram, which is, to be precise, a sine curve, as shown in *Fig.* 4. (This visual tranfor-

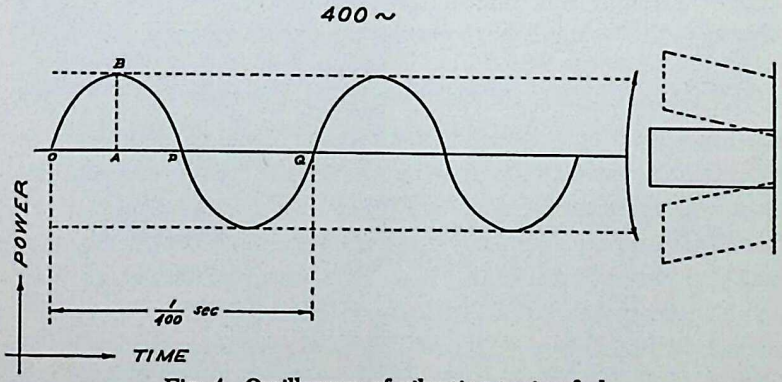


Fig. 4. Oscillogram of vibrating tuning fork.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

mation of vibrations into a curve is merely one of several possible graphic tricks. All such graphs are conventional representations and not actual 'pictures' of a sound 'wave'.) In this diagram the distance AB represents the amplitude (loudness, power) of the tone. The line OQ gives one full cycle, or period. If, let us say, each of the two cycles shown on the illustration is drawn by the stylus in 1/400 of a second, the wave illustrates a tone of 400~. Or, conversely, one cycle lasts 1/400 of a second. In other words, the tuning fork here used has a frequency of 400~, slightly higher than the 396~ of the G above Middle C. The standard A of concert pitch is 435~ in Europe (as adopted by a convention held in Vienna in 1889), and 440~ in the United States. (Note that these pitches of music are conventional, cultural, and not *a priori*, natural standards. We do not know the absolute pitch at which a piece of Bach or Beethoven, to say nothing of an earlier work written in a different notation, sounded when it was first performed, although, of course, all relative pitches, the course of the themes and melodies, are prescribed by the musical notation of the composer and therefore unchangeable. This relativity of pitch which does not affect the recognizability of a piece in musical performance should be kept in mind since it is of importance also in linguistic performance.)

From the frequency of a tone one can reckon its wave length, that is, the length of one period, measured in feet, by dividing the speed of sound per second by the frequency. A 400~ wave has therefore a length of $\frac{1130 \text{ ft}}{400} = 2{,}8$ ft. (On an illustration like Fig. 4 it will not be convenient, or necessary, to represent a wave in its true length. To do this would require, in our experiment leading to Fig. 4, to move the paper at the speed of sound. The velocity of sound, unlike that of light, is not completely invariable. 1130 ft per second is its speed in air at a temperature of 20°C, or 68°F. At 0°C, or 32°F, it is reduced to 1089 ft per second. The velocity increases with the density of the medium: solids are better conductors than liquids,

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

and liquids better ones than gases. Sound cannot traverse a vacuum, while light can.)

Among vibrating bodies the tuning fork occupies a special position since it produces a very simple tone, that is to say, a tone that is made up of chiefly one frequency. (The ocarina and human whistling share this peculiarity.) Because of the simplicity and purity of its tone the tuning fork is employed to provide a reference pitch to piano tuners, *a capella* choirs (singing without accompaniment), and musicians who need to tune their instruments. (Some persons, but relatively few, are endowed with absolute pitch, which enables them to create or recognize with precision a tone of a given frequency without listening to a reference or standard tone.)

But most of the sounds that reach our ear, whether they originate from a musical instrument or the human voice or any odd object, whether they be tones or noises, are not so simply constituted. Instead, they are composed of two or more, often very numerous frequencies, all of which are by nature fused in such a way as to produce the perception of a single sound. Such a sound is called complex.

Since the perception of a complex sound is that of one and not of several sounds, it is to be expected that the oscillogram of a vibrating body producing a complex sound will not consist of several simple sine waves but of a single more complex wave. Yet one may also reasonably presume that this complex wave picture is somehow the sum of the component simple sine waves, and is probably capable of an analysis into its component parts acoustically, mathematically, and visually. Let us examine this proposition in some detail.

To approximate, though not equal, the conditions of a single vibrating body which produces sound at two frequencies (which is still a very simple case), let us strike two tuning forks of differing frequencies and trace their wave picture, or oscillogram. The results are shown in *Fig. 5*. Wave A produced by one tuning fork, has a frequency of 200~, wave B produced by the other tuning fork, of 600~. But what is wave C in this

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

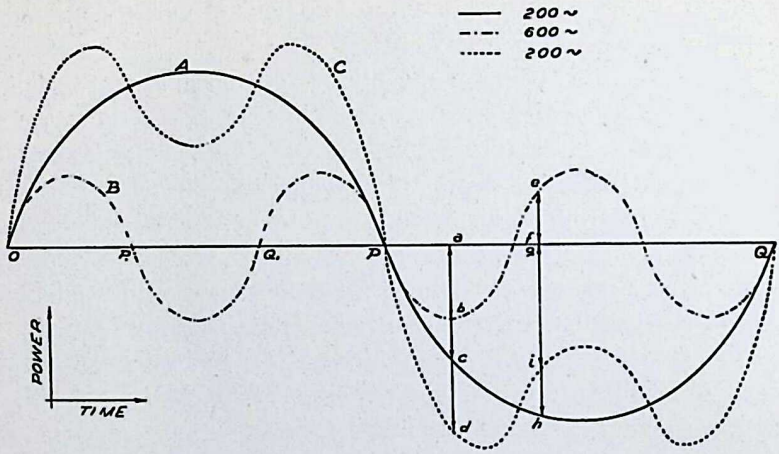


Fig. 5. Oscillogram of two tuning forks.

illustration? It is not a simple sine wave, and therefore does not represent a simple tone, hence is not produced by a third tuning fork. But it has a frequency of 200~, which means that it does sound at the same pitch as wave A. Do waves A and C therefore visualize, despite their different aspect, the same tone, or do they not?

Wave C is, in fact, the sum, the blended result of waves A and B. Mathematically, through a long and very complicated computation, it is possible to determine that wave C is actually the combination of a wave of 200~ and one of 600~. In the oscillogram of Fig. 5 one can construct wace C geometrically by adding waves A and B, thus: ab+ac=ad, or ef+gh=gi, and so forth, at any chosen point. The pitch of the audible tone represented by wave C is 200~, like that of wave A. This is in accordance with an acoustic law which states that the frequency of the perceived pitch of a tone equals the greatest common divisor of all the frequencies actually present in this tone. Since the greatest common divisor of 200 and 600 is 200, the pitch of wave C is also acoustically determined as 200~. We must therefore ask ourselves whether in this fusion

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

of A and B, which results in C, the wave B simply disappears without a trace. In terms of pitch, at least, this is true. On the other hand, the profiles of waves A and C are so different from one another that one cannot help wondering whether, apart from pitch, they really represent the same tone. The answer is that they do not, and that wave B has something to do with the formation of wave C, something through which C becomes, outside of pitch, different from A. Let us see what B's contribution is and in what respect A and C differ acoustically.

Pitch, in particular fundamental pitch, which is the frequency at which a complex tone sounds, is only one component of the tone, but the decisive one as far as melody is concerned. Everyone knows, however, that sameness in pitch of tones does not imply sameness of auditory effects outside of pitch. Hence 'same' tones sound 'different' if played by various instruments or sung by different voices. Were it otherwise, there would be no purpose in employing an orchestra in which a number of different instruments may play exactly the same melody, nor would it make any difference whether a tune were sung by one or another singer. If pitch were the only characteristic of musical performance, all singers who could carry a melody correctly would be equally good performers and their voices would be indistinguishable in the performance of a prescribed melody; and within an orchestra it would not matter whether the same melody were played in unison by a violin, a clarinet, a viola, and a French horn, or just by four violins. It remains to determine what factors account for the variety of instrumental tones and voices performing at the same pitch.

According to Fig. 5 the difference between A and C is the shape of the wave. One says, therefore, that, oscillographically speaking, the PROFILE of the wave is indicative of some non-pitch characteristic. Acoustically this characteristic is called the QUALITY of a tone. Now the contribution of wave B in the formation of wave C is obvious: it affects not the length, but the profile of each cycle of the wave, that is, not the pitch, but the quality of the resultant complex tone. In other words,

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

A and C might represent tones of like pitch played by different instruments or sung by different singers. (I disregard here the fact that most musical instruments and all singing voices actually produce considerably more complex tones.)

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

III

OSCILLOGRAMS

Wave C is a synthesis of waves A and B, and we can calculate its acoustical characteristics by adding its components. Through the analogous computation in reverse we can decompose wave C into waves A and B (which, we know, must be sine curves). This process is called HARMONIC ANALYSIS. The purpose of such an analysis is, then, to determine the frequency composition of a complex wave. The result of harmonic analysis delivers a statement on the frequency of the FUNDAMENTAL, that is, on the pitch at which a wave is perceived, and on the number and frequencies of OVERTONES, or HARMONICS, present in the tone. (In acoustic terminology, the fundamental is the first harmonic; the second harmonic is the first overtone, the third harmonic the second overtone, etc.) Overtones occur always in integral multiples of the frequency of the fundamental. By implication they are weaker than the fundamental, but their power does not necessarily decrease in inverse ratio to their distance in cycles from the fundamental. Just as graphically the profile of the complex wave is determined not only by the presence of the various frequencies but also by their amplitudes, so the quality of the tone depends also on the distribution of power among the frequencies. The configuration of these powers is, of course, an inherent quality of the tone and therefore due to the physical peculiarities and the manner of excitation of the vibrating body which produces it. But this primary tone (produced, for example, by the vocal lips, a violin string, the reed of a clarinet) may be further modified, before it reaches our ear, by its passage through resonance and filter chambers and by reinforcement (for example, in the oral and nasal cavities, the body of a violin, the volutions of a trumpet). I shall speak of these phenomena in their proper place.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

If in Fig. 5 one added further overtones at multiples of 200, say, at 400~, 800~, 1000~, and 1200~, the resultant tone would necessarily still sound at a fundamental frequency of 200~, this still being the greatest common divisor of the frequencies present, but its quality would be quite different from that represented by the original wave C because of its altered harmonic composition. Naturally, the oscillogram resulting from adding the new set of component frequencies would no longer resemble wave C.

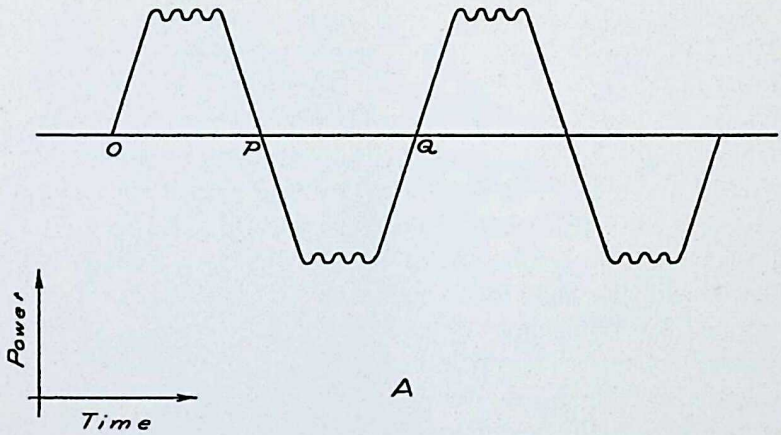
But if we should discover, by harmonic analysis, that a certain tone contains harmonics not only at 400~, 600~, 800~, 1000~, and 1200~, but also one at 300~, we should know at once that, the common divisor now being 100, the fundamental frequency of this complex tone is 100~, or one octave lower than the one of 200~. It should be carefully noted here that the acoustically perceived frequency of 100~ is not present among the component frequencies. A complex tone may accordingly sound at a frequency undiscoverable by harmonic analysis. In other words, the so-called fundamental, which is generally responsible for the audible pitch of a complex tone, need not be among its components.

Are *all* waves reducible by means of harmonic analysis? Yes —*provided* that they are exactly repetitive, or periodic, waves. This means that the period OQ on my diagrams has to recur over a period of time in the same dimension and in the same oscillographic shape, or profile. A sound represented thus by a periodic wave is called a tone, and distinguished from a noise which cannot be so represented because it is composed, at least in part, of random frequencies. This is the physical reason why a tone is a sound of measurable pitch, whereas a noise is not.

Also the so-called square waves of *Fig. 6* are actually sums of sine waves. Fig. 6A represents the sum of few, four or five, components; the more of them are added, the more the resultant complex wave comes to resemble Fig. 6B. (To obtain the square wave oscillograms shown, only the odd harmonics, 3rd, 5th, 7th etc., are used.) The repetitiveness alone of these shapes

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri



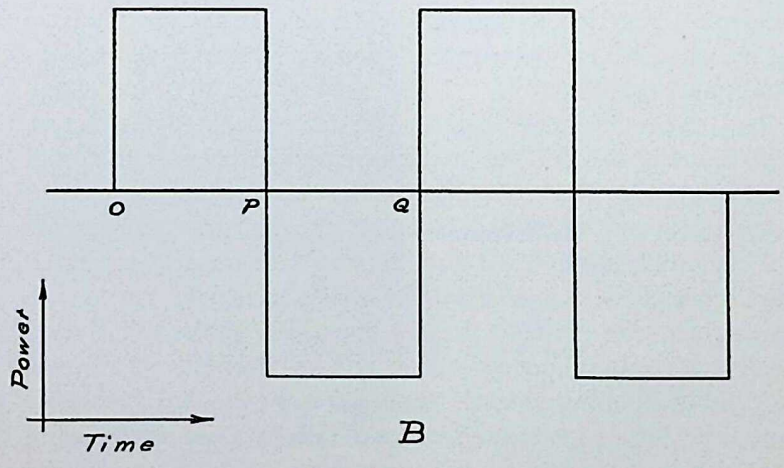


Fig. 6. Square waves.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

vouches for their susceptibility of analysis into component sine waves.

Now an ideally repetitive wave would of course have to last in the atmosphere into eternity, since it contains no internal force to change it or stop it once it has been set in motion. (Under such circumstances one could imagine oneself recapturing a sound wave which has been travelling for millennia. An overheard oration of Demosthenes, or a conversation in a tavern of ancient Pompeii would be of enormous interest to the historical linguist!) In the physical world, however, conditions are never mathematically ideal, and sound waves come to an end through loss of energy inflicted by innumerable more or less yielding obstacles, of which the atmosphere itself is an ubiquitous and inevitable one. Yet air is only incidentally an obstacle; primarily it is the sound's indispensable vehicle, because, unlike light, sound cannot traverse airless space. But the change and decrease in the power of sounds in speech communication are gradual, no matter how rapid, and for the practical purposes of acoustics of speech sounds and phonetics we may take a slice out of a finite wave and deal with it mathematically as though it were ideally infinite and non-diminishing. This is, in fact, what I have done in the preceding examples and illustrations.

All the oscillograms discussed so far have had one feature in common, in addition to their being repetitive, namely, all the component waves constituting the complex tone start at the same moment, that is, at a common zero point on the time axis of the oscillogram. Since, representing harmonics, they are all multiples of the fundamental in frequency, they all must reach a common zero point on the baseline again, at the end of each half-cycle of the fundamental (regardless of whether the fundamental is actually part of the tone and oscillographically present or not). Consequently the second half-cycle of the fundamental not only is a negative mirror picture of the first but also encompasses a repetition of all the partial waves occurring in its first half; *ipso facto* it is also a repetition of the complex wave

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

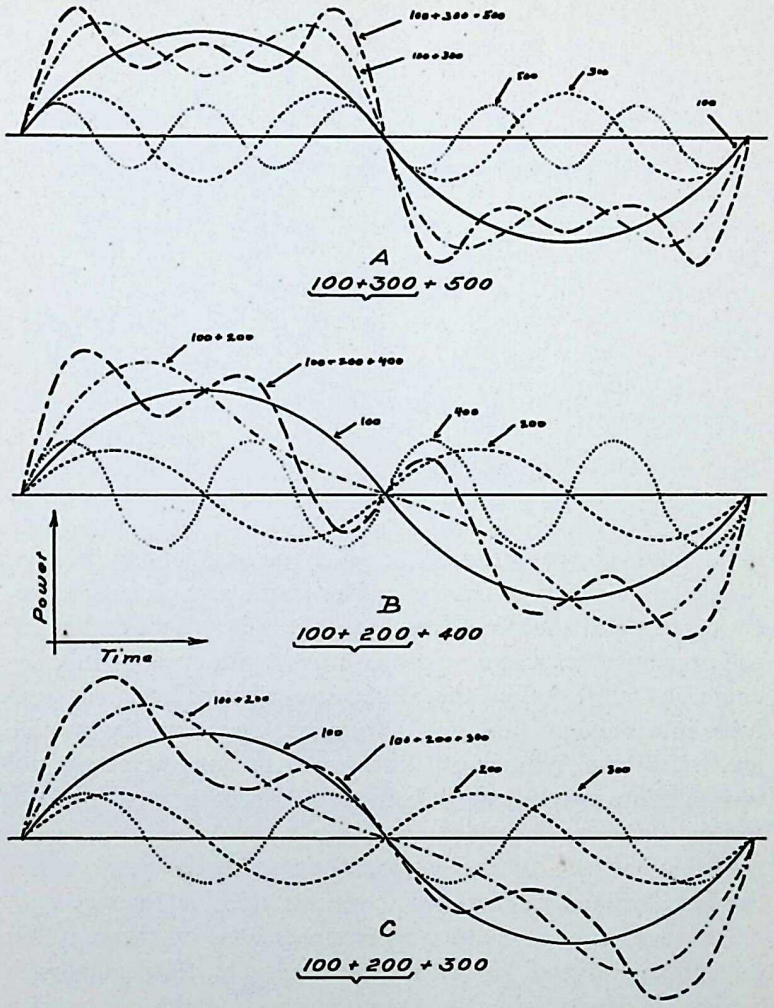


Fig. 7. Complex wave—symmetrical.

produced by all of them together. This is illustrated in *Fig.* 7, where various combinations of the fundamental with odd, even, and mixed harmonics together with their sums are shown.

We may now compose a complex wave of which no two components meet on the zero line. It is illustrated in *Fig.* 8.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

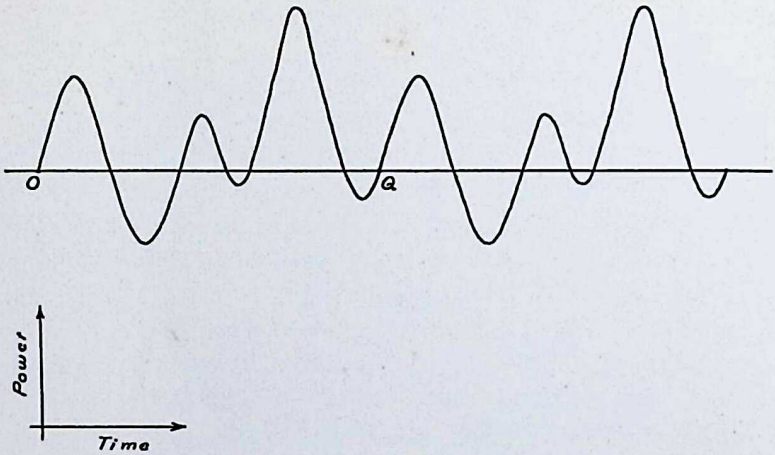


Fig. 8. Complex wave—non-symmetrical.

(The physical requirements for such an oscillogram will be discussed in the next chapter.) The significant difference between this and the preceding shapes is that, although like all the previous ones it is a repetitive wave (with cycle boundaries again at O and Q) and therefore susceptible of harmonic analysis into fundamental and overtones, and also to graphic analysis into a number of sine curves, its full cycles are not divisible into two mirroring half-cycles. In other words, unlike the preceding waves, this one does not have a point P on the zero line halfway between O and Q at which the wave crosses the baseline and after which it repeats itself in reverse.

An extreme case of this type is represented by the so-called rectangular wave of *Fig. 9*. We are obviously still dealing with a periodic repetitive wave (note the period O-Q), hence with a tone which can be subjected to harmonic analysis. I call it an extreme case because the part of the wave above the baseline lasts only 1/10 of the period, the rest 9/10. (My drawing is not in scale.) To obtain this peculiar profile we need an arrangement of harmonics and powers (amplitudes) as described in *Table II*. This table shows that, if the power of the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

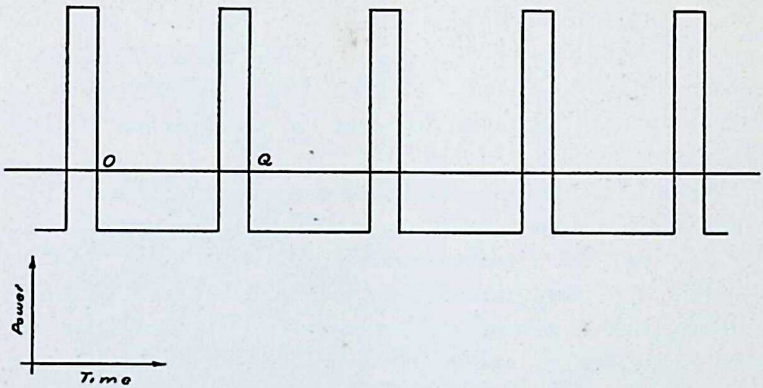


Fig. 9. Rectangular wave.

first harmonic, or fundamental, is fixed as 1.000, the sixth harmonic is still more than half as intense as the first. This means, of course, that at least the lowest five overtones, though naturally they have nothing to do with the pitch, do contribute a great deal to the quality of the tone, or, graphically speaking,

Table II

POWER DISTRIBUTION IN HARMONICS OF RECTANGULAR WAVE

Harmonic	Amplitude	Harmonic	Amplitude
1	1.000	16	0.192
2	0.951	17	0.154
3	0.873	18	0.106
4	0.770	19	0.053
5	0.647	20	0.000
6	0.513	21	0.048
7	0.374	22	0.086
8	0.238	23	0.114
9	0.111	24	0.128
10	0.000	25	0.129
11	0.091	26	0.118
12	0.159	27	0.097
13	0.201	28	0.068
14	0.220	29	0.034
15	0.216	30	0.000
		etc.	etc.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

to the profile of the wave. Note also that in this particular arrangement the 10th, 20th, and 30th harmonics are missing, as would be, if our table went high enough, the 40th, 50th, etc. The number of harmonics present is, as in Fig. 6, theoretically infinite.

When I just said that the overtones have nothing to do with the pitch of a tone, I was speaking strictly in terms of physics, in cps. But if I tranferred my statement to the level of perception, expressing myself in mels, I should have to say that different overtone composition may indeed cause different perception of pitch, even though the frequency of the fundamental remains the same. Again we are brought face to face with the necessity of viewing acoustic events with proper regard to the psycho-acoustic reality of the listener's perception.

For analogous reasons, when I spoke of adding tones and obtaining thereby a complex tone that has the greatest common divisor of the components as its fundamental frequency, I was correct mathematically; but in order to attain audible results that bear out my acoustic statement, in order actually to hear a tone of the predicted pitch, I must see to it that the various fundamentals, with which I reckoned without paying much attention to the overtone situation, be actually accompanied by a great number of upper harmonics occurring with an appropriate distribution of energy. Otherwise mathematics and perception will not agree, that is, the mels perceived will not correspond to the cps calculated.

In the description of speech sounds, however, the possible disparity among mel and cps statements is of small consequence; first, because absolute pitch is unimportant in speech; second, because speech sounds of measurable pitch or frequency do in any event have numerous and powerful overtones.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

IV

GLOTTAL TONE

I have led the discussion toward oscillograms representing tones of highest complexity because the tone produced by the vocal lips, the glottal tone, is, most scientists agree, of that kind. It has been suggested that the oscillogram of the glottal tone comes near to being the rectangular wave of Fig. 9; but it is also admitted that this oscillogram, together with its harmonic composition (Tab. II), is to a great degree speculative. This lack of sure knowledge is not surprising if one considers that the glottal tone, as emitted by the vocal lips but still unaffected by supra-glottal modulations in the resonance chambers of mouth and nose, is not easily accessible for acoustic investigation. All kinds of insertions of microphones, either through the mouth or into the glottis directly from the outside (in cases of persons having undergone laryngectomy) have been tried, but all make for unnatural tone production.

The theory that each vibration of the vocal lips is due to a separate innervation and that, consequently, a nervous center directly controls the frequency (pitch) of a tone, has been largely abandoned. Instead, it seems to be proven that the glottal frequency results from the combined effects of subglottic pressure and the mass and tension of the vocal lips, with variations and changes in frequency controled by a feedback, that is, by the performer listening to himself and using his kinaesthetic sense for regulation.

Excellent slow-motion (4000 frames per second) films of the vocal lips in action have been made, and many experiments have been performed, by means of catheters in the glottis and the oesophagus, to measure subglottic pressure during phonation. A good deal has been learned, therefore, on the kinds

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

of movements which the vocal lips execute, on their reactions to variation in sub-glottic and supra-glottic pressure, on the relation of vibration to voice quality, etc. Yet scientists are, at this moment, still far from knowing everything that needs to be known about the glottal tone.

For our present purpose, fortunately, ignorance of many acoustic data pertaining to the glottal tone is, no matter how regrettable, not fatal. We may, for the moment, be satisfied with the knowledge of the one detail on which acousticians seem to agree, although full experimental proof and mathematical calculations are still lacking. This is the view that the glottal tone is one of very great complexity, that it consists of a vast number of harmonics, and that many of these harmonics are strong and therefore contribute significantly to the quality of the glottal tone.

It is also known that not all harmonics composing the glottal tone commence at the same moment. This is due to the inclined position of the vocal lips in relation to the direction of the air stream that activates them, causing the lower portions of the vocal lips to begin vibrating before the higher portions. If an oscillogram of the glottal tone were therefore drawn, it would resemble Figs. 8 and 9 in that its half-cycles do not mirror one another.

It is to be expected that a tone already so rich in powerful harmonics can easily be modulated in quality and power through the effects of resonance, filtering, and reinforcement. These phenomena, which will be dealt with in detail, are caused by the shape and size of the resonance chambers through which the tone passes before it reaches the listener's ear. In the case of the human vocal apparatus, the supra-glottal cavities fulfill this function. These cavities can be quickly and nimbly altered so as to resonate at, or filter out, or reinforce, certain frequencies contained in the glottal tone, thanks to the great mobility of the organs constituting them.

The composition of the glottal tone and the constitution of the speech organs are such that small dislocations of the organs

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

suffice to produce perceptible and significant acoustic variations in terms of both pitch and quality. The number of possible variations is, therefore, enormous. But from this vast total each language selects but a small group of distinct classes, or phonemes—voiced phonemes, as long as they contain the glottal tone.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

V

PHASE

I said that in cases where not all components of a complex tone start at the same zero point, that is, at the same time, one obtains oscillograms of repetitive waves whose half-cycles are not mirror pictures of one another. This leads us to the consideration of a feature of oscillography and acoustics neglected so far, that of PHASE.

That waves A and B of Fig. 5 begin to vibrate at the same time was a possible but gratuitous assumption. Reality can be like this, but need not be. Indeed I have just mentioned, in Chapter IV, that the glottal tone especially is one in which this condition does not prevail. If we repeat the experiment of Fig. 5 but, instead of activating both tuning forks at the same time, delay the striking of the second fork by a fraction of a second smaller than the completion of a half-cycle of the first fork, that is, if we begin on our sketch the oscillogram of B at a point between the beginning and the end of a half-cycle of A, we now obtain by adding A and B a wave C_1, as shown on *Fig. 10*, which is obviously and necessarily different from wave C of Fig. 5. We have changed, not the component waves, but only the phase. The question is now whether waves C and C_1 which differ in phase and profile, but in nothing else, sound alike or not. Or, to put it in other terms, does variation in phase, if all other factors remain equal, cause a difference not only in the oscillogram but also in the audible perception?

The answer, which is easily discovered by experimentation, is that the tones represented by C and C_1 sound in fact wholly alike, in both pitch and quality. (Only under certain special conditions, which need not concern us here, may phase have an influence on a tone. In fact, if there should ever be two

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

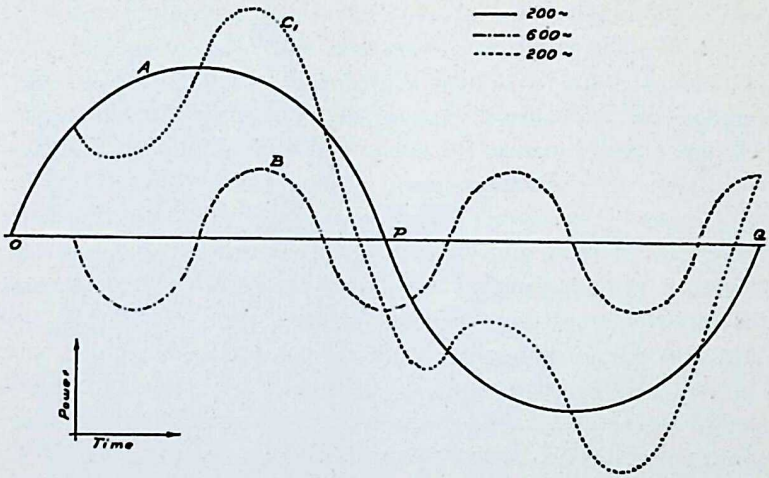


Fig. 10. Complex wave—non-symmetrical.

exactly equal components which operate at exactly opposite phases, that is, with the second beginning at the end of the half-cycle of the first, they would cancel one another not only oscillographically but also acoustically. The result would be, strange as it seems, silence.) Consequently, variations in oscillographic shapes, or profiles, at the same frequency, may represent not only acoustically perceptible variations of quality but also acoustically non-perceptible variations of phase.

Evidently, while our ear and auditory nerve centers transmit differences in the composition of a complex wave in frequency and amplitude, or pitch and loudness, of fundamental and overtones, the phase is not perceived or reported. When the human brain subjects an audible impression to a process equivalent to a harmonic analysis of the airwaves that reach the ear, and reports the pitch and the quality of the tone, it automatically makes those adjustments and corrections which neutralize the phase and render it imperceptible.

Does this mean, then, that an oscillogram is a false visualization of reality? No, it merely means that the acoustic reality

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

of the sound which it represents does not agree with the acoustic reality which we perceive. In other words, the oscillograph is a machine which, while it delivers no wrong acoustic information, does not furnish us with the kind of information which is most significant in terms of normal human aural perception. And since it is our task here to examine how sounds are perceived, and what sounds are perceived, we need a machine that gives information which is pertinent in this respect. On the basis of these premises, the phase-determined oscillogram is therefore a graph that registers optically the aurally irrelevant physical phenomenon of phase. The physicist may well find himself in a position where he wishes to have detailed knowledge on the phase. But the linguist will scarcely care; he will, on the contrary, want to be assured that any picture of a sound he looks at is no more determined or influenced by phase than is the aural and nervous perception and identification of a sound, which is not at all.

Hence an acoustic description of a speech sound satisfactory to the linguist (*qua* linguist) need contain less information than is furnished by an accurately drawn oscillogram, the superfluous quantity being the phase. All that our purposes require for a complete description of a complex sound is a statement of how much power is present, at a given moment, at every frequency within the range audible to the normal human ear, which is from about 20~ to 20,000~ (in children up to 40,000~). (But within this range it is mostly unnecessary to go higher, for linguistic purposes, than about 8000~, and even 3600~ has proved, by and large, a satisfactory upper limit, which excludes only some typical high random frequencies of fricative sounds. Hence few of the spectrograms I shall subsequently discuss go beyond 3600~.) Now such a statement on the distribution of power is rendered, for example, by Table II above.

But my ultimate aim is to arrive at a visualization of speech sounds, indeed at their registration as a function of time, so that they may be represented as nearly as possible in the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

manner in which they occur in the actual stream of speech. Hence Tables of the type shown will not serve since they list separately and successively the components of the sounds, or those components of each sound which one takes the time to calculate at any number of points in its total duration.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

VI

SPECTRUM

One manner of visualization of a complex wave is given in *Fig. 11.* In order not to make its reading too complicated, I

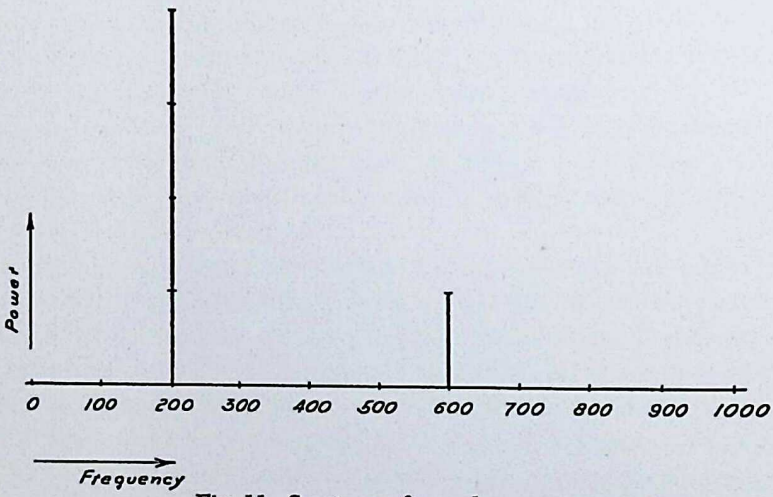


Fig. 11. Spectrum of complex wave.

have chosen to present in it the situation of Fig. 5 (two components), rather than that of Fig. 9 (infinite number of components). We know that wave C of Fig. 5 is composed of waves A and B and therefore has power at the 200~ and 600~ levels. The amount of power is indicated in the amplitude of the wave on the sketch. Mathematically, acoustics measures power in terms of the square of the amplitude. If we assume, then, that the amplitude of the 200~ wave is 2, and that of the 600~ wave is 1, we obtain at 200~ a power of 4, and at 600~ a power of 1. Hence on the diagram of

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Fig. 11, which shows cycles on the horizontal (abscissa) and power on the vertical (ordinate), the two vertical lines indicate the amount of power present at the frequencies of which wave C is composed. If now we perform the same operation for wave C_1 of Fig. 10, we obtain, since only amplitude and frequency but no longer the phase are represented, exactly the same result. Fig. 11 represents both waves C and C_1, that is to say, like perception is now rendered by like graphs. Hence this may be considered a diagram which is truer to the audible reality perceived by the listener than is an oscillogram, providing us only with relevant information on the amount of power at the frequencies present, and omitting the irrelevant information on phase. Fig. 11 is called the SPECTRUM of a sound.

However, in the process of constructing this spectrum we have lost the element of time, which was present in the oscillogram. The reason is, of course, that on a two-dimensional representation we can render only two dimensions of a sound, either a time and frequency profile, or a power and frequency profile. This kind of spectrum, therefore, does not by itself tell us whether it describes completely a (theoretically) endless or continuous tone which throughout its life operates at these two frequencies with these two powers, or whether it is a fortuitous snapshot representing a momentary acoustic reality within a series of successive tones, staccato or gliding, recorded accidentally at a moment when this particular configuration of frequencies and powers prevailed.

We could construct a three-dimensional analogue so as to retain also the third dimension, time. But since our goal is a machine which records sound graphically on a plane rather than in space, we shall have to be satisfied with a two-dimensional schema. Obviously, the parameter of time will be of the essence in a succession of sounds which change as rapidly as do those of the stream of speech. Hence we shall have to forego an exact recording of one of the other two parameters, frequency or power. Here we shall, as linguists,

डा० विद्या निवास मिश्र
प्राध्यापक, भाषा विज्ञान
निजी पुस्तकालय
पुस्तक संख्या

Vidya Niwas Misra
Linguistics Department
University of Gorakhpur,
GORAKHPUR.

Vidya Niwas Misra Collection, Varanasi. Courtesy Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

prefer to neglect a notation of absolute (but not relative) power rather than frequency. First, because the absolute scale of loudness is a non-distinctive quantity in language—unless there be a language in which a shouted sequence of phonemes, or a word, has a lexical meaning different from that of the same sequence pronounced at a lower volume. (Such a language would make it impossible, for example, to raise or lower the volume of radios, and it would necessitate that telephones produce speech at the natural level of loudness; to say nothing of the insurmountable difficulty of obliging all speakers to make same utterances at the same level of loudness, and of the ridiculous situation wherein the inevitably unequal acuity of hearing of several persons would make it impossible for a speaker to make himself understood by all at the same time. Furthermore, a speaker would have to adjust the loudness of his utterance to the distance at which the hearer finds himself, because it is the loudness of perception and not of emission that determines intelligibility; or else the hearer would have to recalculate what he hears in terms of this distance.) Second, because relative loudness, so-called stress, is only relationally distinctive within a sequence, so that the question is never how much power is present, but how much more or less power is present in one segment as compared with another. And the relative distribution of power among the frequencies present can be marked on our visualization even without introducing a third dimension, as we shall see.

All visualizations of sound in general and of speech sounds in particular were until recently of the oscillographic type, from the old kymograph, with its stylus activated by a membrane and scratching lines upon a sooted scroll of paper which was unrolled in the manner represented by Fig. 4, to the sensitive cathode ray oscilloscope. (It should be noted, however, that the kymograph produces registrations representing variations in the total amount of pressure during articulation, while the oscilloscopic vibrations trace single air particles). From all these representations it is difficult to abstract the relevant linguistic

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

information because much linguistically irrelevant material is incorporated in their fabrication. In addition to phase, individual voice pitches, too, play a large part, which are of course also non-significant in language since it makes no difference whether a speech utterance is made by a high-pitched or by a low-pitched voice, by a man, a woman, or a child. Relative variation of pitch may be distinctive on the lexical level in so-called tone languages, but it is, like relative stress, of no consequence in phonemic segmentation. Relative pitch in non-tone languages is intonation, and it, too, has nothing to do with phonemic segmentation. (If an utterance in a tone language is also provided with some intonation pattern, then the total audible pitch contour is the sum of tone plus intonation, plus the inevitable tone sandhi, which is the influence of a neighboring tone upon the tone under examination—cf. sandhi in comparative grammar.)

With some training one can learn to recognize recurrent types of oscillographic shapes (which one might call oscillophonemes) and associate them with certain speech sounds (phonemes) despite these handicaps, but the task is both difficult and hazardous. A more serious draw-back is the impossibility of performing a great number of linguistically significant experiments by means of oscillographic machines, and of making accurate measurements of certain linguistically significant parameters of speech, which for this purpose must not be distorted or falsified in their graphic representation, as in fact they are in oscillograms, by the intrusion of non-significant factors. Hence that kind of apparatus is, in the hands of linguists, chiefly suited to demonstration but not to quantifying investigation. The oscillograms support the impressionistically gathered information on (speech-)sounds, but they deliver considerably less than all linguistically relevant and accessible data. The machine which I shall describe presently supplies a graph containing the information provided by a spectrum (see Fig. 11), hence is called the sound spectrograph, by performing a harmonic analysis in many ways analogous to that

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

performed by the human hearing mechanism and its nervous centers. But before passing on to that, a few preparatory technical remarks are required.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

VII

RESONANCE, REINFORCEMENT

Suppose I have in a room three silent tuning forks whose natural pitches (that is, the frequencies at which they operate when struck) are 200~, 300~, and 400~. If I then strike a fourth tuning fork of 300~, it will not only sound itself, but it will also communicate its vibrations to the hitherto quiet fork of 300~, while the other two remain unaffected. This effect of sympathetic vibration is called RESONANCE, and the object activated by it is a resonator.

There may well be other objects and surfaces and airspaces in the room with a natural frequency of 300~, and they too are set to vibrate. But their responsiveness and sensitivity, unlike those of the tuning fork, are likely to be so small that the amplitude of their sympathetic vibrations will not evoke a sound of sufficient loudness to be perceptible to the human ear, nor one of perceptible duration.

The resonator not only becomes active shortly after the sonator, but its motion, like that of the pendulum, continues for a while on stored energy, even after the primary source ceases to deliver energy. Hence the resonating tuning fork continues to operate for a while after the struck fork has been stopped. But it always functions at a lower level of power than the original source of sound, the reduction being due to loss in transit and in the instrument itself, mainly through friction in which energy is dissipated as heat.

Resonance must not be confused with reverberation, or echo. This is a reflexion, a bouncing back of the original sound waves from a non-absorbent surface, and not the activation of an additional vibrator. In reverberation no storing of energy takes place and the duration of the sound is not extended beyond the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

operation of its source, except to the extent that the listener's ear is struck by echoing waves after a lag of time proportionate to the distance they had to travel from the source to the reflecting surface and thence to the ear. Since the non-reflected waves reach the hearer more quickly, he actually perceives the sound twice, or even more often if there is a multiple echo. A certain amount of reverberation is indeed sought in concert halls and opera houses, and contributes to the good acoustics of a room. It can be eliminated by clothing the walls of a room with absorbent, 'deadening' material, as is often done in radio and recording studios, or in laboratories where acoustic experiments take place.

Let us return to our resonating tuning forks. One may ask next whether also a resonator of 600~ would have responded to a 300~ tuning fork. I mentioned before that the peculiarity of the tuning fork which makes it attractive to musicians desiring to tune their instruments, is, in addition to the long duration of the tone without precipitate weakening after a single striking, the great strength of the fundamental and the virtual absence, apart from a few faint upper harmonics, of important qualifying overtones. Hence a tuning fork vibrating at 300~ would produce so weak a harmonic at 600~, if any, that a resonator of 600~, struck by very small acoustic power, might not respond audibly. By the same token, an ideal tuning fork can resonate at its own fundamental frequency only, and even a less than ideal tuning fork will not be an efficient resonator at frequencies other than its fundamental. But if the same experiment is performed with a vibrating body rich in strong overtones, a violin string, for example, then chances are that a 300~ string will give forth vigorous enough a harmonic at 600~ so as to cause a sufficiently sensitive 600~ resonator to react.

Experiments show, then, that a body, or a volume of air enclosed in a cavity of a certain size and shape (in a bottle, or in a trumpet, or in the cavities of mouth and nose), can be caused to vibrate and to produce a sound at its natural frequency

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

not only if directly activated by an external source of power, but also if merely exposed to sound waves operating at its frequency, provided they occur with sufficient energy. In other words, it can be caused not only to sound but also to re-sound, to be, in short, a resonator.

If we imagine the very complex glottal tone entering the cavities of the supra-glottal speech organs, it is evident that sufficiently powerful harmonics may cause suitably tuned air spaces which they encounter to resound fairly vigorously. Now the glottal tone has a vast number of harmonics, and many of them are very powerful. On the other hand, the air spaces of the cavities can be made to assume a vast variety of shapes and sizes thanks to the mobility of the vocal organs, and the configuration of these air spaces can be altered rapidly and at will by the speaker. Therefore, the air spaces of the cavities can be arranged so as to acquire, in rapid succession, a great many different inherent frequencies at which they may resonate. The activity of the speaker consists in matching the resonating cavities with one or several harmonics of the glottal tone, that is to say, in creating cavities of a size and shape that will resonate if activated by one or several of the harmonics contained in the glottal tone. Every resonance thus originated and added to the glottal tone changes the oscillogram, the profile, the spectrum, in short, the quality of the glottal tone. Consequently, the tone which eventually issues from the mouth may be considerably and significantly altered in quality by the supra-glottal modifications, or, in terms of articulation, by the position of the vocal organs. Most conspicuously, the different vowels of a language owe their characteristic quality, regardless of pitch, to these supra-glottal articulations.

In addition, these resonated frequencies are also reinforced within the supra-glottal cavities. To explain this phenomenon of REINFORCEMENT let us experiment once more with a struck tuning fork. By placing its stem upon a hollow box we may be able to increase the loudness at which it operates, to reinforce its power. This gain in amplitude is obtained in exchange for

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

loss of duration, for a tuning fork thus placed for reinforcement will cease to operate more quickly than one not so placed. It cannot be otherwise: the increase in amplitude must be due to some shift in use of the total energy input. Since the energy input equals the energy transmitted to the fork by striking, and since thereafter no further energy is added, no single part of the total energy output can be increased except by reduction of another part. In other words, the total energy output cannot be greater than the input.

Now the supra-glottal cavities represent such reinforcement chambers (as do the bellies of drums, the bodies of violins, etc.) in which the amplitude of a sounding or resounding frequency is reinforced. The reduction of duration is, in speaking and singing, as in played instruments, of no consequence since the input of energy due to the volitional muscular activity of the performer can be continued and repeated at will. It is as though the tuning fork were repeatedly struck while placed on the box, so that its energy is always renewed before it is completely spent. Consequently, also a frequency of relatively low power in the spectrum of the glottal tone can be made to resonate, thanks to reinforcement, at a higher level of power.

It should be noted that the terms 'resonance' and 'resonance chamber' are often employed so as to pertain indiscriminately or jointly to the phenomena of both resonance and reinforcement. In the present context it is convenient to keep them apart, letting resonance refer to matters of overtone composition, or tone quality, and reinforcement to power, or loudness. Since in the analysis of speech sounds quality is a significant, but (absolute) power a non-significant parameter, features due to resonance will, in the subsequent discussion, be vastly more important than those of reinforcement.

Thus the speaker of a language is operating an extraordinarily sensitive and complex apparatus, by means of which he is enabled to produce an enormously large variety of acoustic, phonetic performances. I have already mentioned that from this potential a given language selects and utilizes only a cer-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

tain number of classes, or phonemes. Playing this 'instrument' is very difficult indeed, or rather it would be if the speaker were obliged to perform all his articulatory actions consciously and laboriously. Instead, he learns during the years of his childhood how to impart to his vocal organs a number of set combinations of articulatory positions productive of the classes of sounds that are the phonemes of his native tongue. These articulations he learns so well and they become so much a part of his behavior that he employs them automatically and without reflexion, and that the acquisition of a new and conflicting set of phonemes, those of a foreign language, becomes a very onerous task, much more onerous for most persons over 12 years of age, in fact, than the acquisition of non-linguistic new sets of classes of behavior, such as those of dressing, or eating according to new rules.

Not all sounds of a language are produced, either exclusively or partly, through the modulation of the glottal tone in the supra-glottal cavities. When there is no glottal tone, other, though no less complex, articulatory activities create the sound. The ones that are produced, wholly or in part, by means of supra-glottal resonances form a class by themselves and are called resonants (as distinguished from the non-resonant fricatives and plosives), and include what are commonly termed vowels (semi-vowels), nasal continuants, and trills and laterals.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

VIII

DAMPING

Unlike the ideal tuning fork, however, the sonators and resonators of the speech apparatus, and also most 'natural' sonators and resonators which surround us, do not produce, and therefore do not resonate at, exclusively the precise frequency of the fundamental and of each overtone. Instead, while the frequency of each of the harmonics constitutes a power peak, it is accompanied by a greater or smaller number of subsidiary frequencies just above and below its own level.

To examine the conditions and consequences of this phenomenon let us convert an ideal tuning fork into one that behaves in this particular manner. This is easily accomplished by squeezing a wad of surgical cotton between the prongs of the fork before striking it. The process is called DAMPING, and the result is a damped sound. This alteration has two significant oscillographic and acoustic effects upon the sound wave now emitted by the tuning fork.

First, damping considerably accelerates the rate of amplitude decrease in the wave, so that the tuning fork may cease to sound audibly after as little as half a second (and thus, apart from the sound quality it now emits, becomes useless as an aid for tuning musical instruments). *Fig. 12* shows three tuning fork oscillograms, A being that of a normal and nearly ideal fork, B that of the same fork if damped, and C if damped even more.

Second, it turns out that damping detracts also from the idealness of the tuning fork with regard to frequencies, in that it originates a number of subsidiary frequencies above and below the fork's natural frequency (which is not shown on Fig. 12). (It does not, however, add new harmonics to the tone a

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

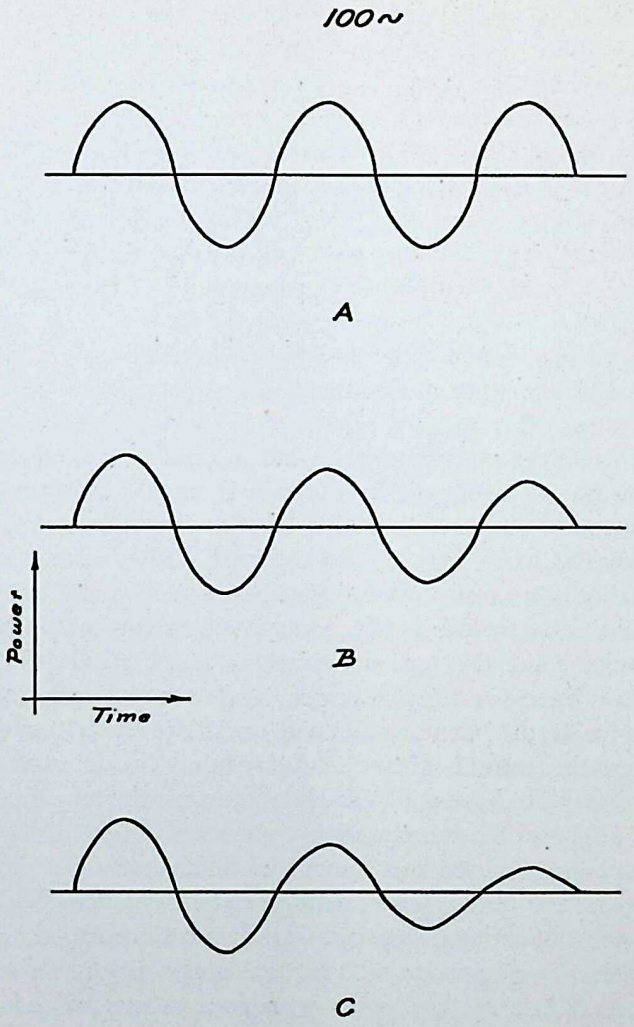


Fig. 12. Oscillogram of damped sound.

vibrating body normally produces.) Theoretically and ideally the subsidiary frequencies may lie anywhere between zero and infinity; practically they occur within a narrower range. These frequencies are clustered around the fundamental (or any over-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

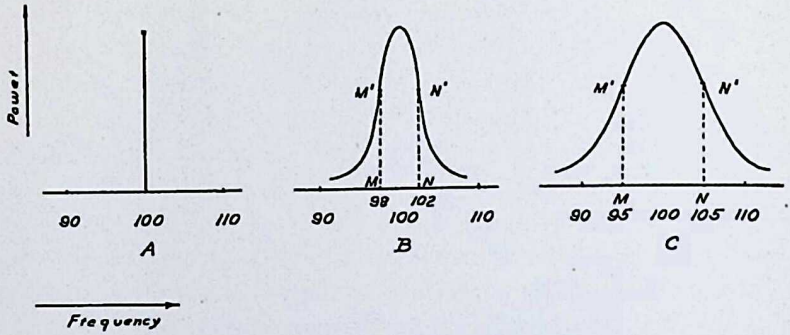


Fig. 13. Spectrum of damped sound.

tones if such are present) of the wave in such a way that those nearest the frequency peak represented by the fundamental (or overtones) are strongest. But their power decreases fairly rapidly with their distance from the peak. Their influence on the quality of a musical tone is generally small, partly because of their weakness, and partly because, not being multiples of the fundamental, they are not overtones. But musical instruments can be so constructed, or provided with such accessories, as to give forth distinctly damped tones. In speech sounds, however, the subsidiary frequencies caused by damping are strong enough to have a noticeable effect upon the tone quality.

Fig. 13 shows the spectra of three tones of different degrees of damping, corresponding (though no mathematical accuracy was attempted) to the three oscillograms of Fig. 12. While the oscillograms illustrated the time-power relationship, showing the decrease in power caused by damping as a function of time, the spectra illustrate the frequency-power relationship, showing the presence of subsidiary frequencies clustering around a peak as a result of damping. The spectrum of Fig. 13A is constructed just like that of Fig. 11, except that it has only one instead of two components, that is to say, power is present at only one frequency, which is what one expects from an ideal tuning fork. In the spectra 13B and 13C there should be above

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

and below the fundamental frequency of the now damped fork an infinite number of subsidiary frequencies which rapidly dwindle in power on either side of 100~. Instead of drawing in all or some of these frequencies as vertical lines, I merely connected their tips by means of an envelope curve, which is generally referred to as RESONANCE CURVE, or SPECTRUM CURVE.

Now the damping effect can be reduced or neutralized, as regards both amplitude and frequency, if the sonator is continuously driven by an external source of power, whereby the decay of the sonator's own power is prevented. Thus, for example, the damping effect of friction upon a clock's pendulum is neutralized by the continued power provided by the clock-spring: as long as the clock is wound, friction will not make it stop. The stream of air issuing from the lungs is such a source of power, so that the effect of damping imposed upon the glottal tone by the anatomic peculiarities of the vocal lips may be eliminated. We may, consequently, consider the original glottal tone as undamped, which is exactly what I have done so far. But the resonating and reinforcing cavities in the mouth, which impart to the speech sound a great share of its perceived acoustic quality, are not so driven by continuous power. Hence completed speech sounds as they issue from the mouth or nose are indeed subject to the effects of damping, as are their oscillograms and spectra. Therefore, our visual representation of speech sounds which we desire to parallel the acoustic form in which the sounds reach a listener's ear, must be so constituted as to register this damping effect, that is, it must show the range or ranges of frequencies in which damping takes place. In other words, the spectrogram must contain information on the effect of damping upon the frequency distribution of a complex tone.

Theoretically the resonance curve of a damped sonator never reaches zero, and theoretically all frequencies between the central peak and zero are present in the sound: the resonance curve approaches asymptotically, but never reaches, a zero level of power. Hence the breadth of each resonance curve is infinite.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

But since the power at frequencies farther removed from the center is infinitesimally small and therefore acoustically negligible, a practical boundary has been set for the effective breadth of such curves. It is measured in cps, and is determined as the distance in frequencies between these two points on the curve on either side of its top where the amplitude is 0.707 times the maximum of the peak. These are called the halfpower points, whose power is, as usually, proportional to the square of the amplitude. (The square of 0.707 equals 0.499849, or just about one-half.) On Fig. 13B these points lie at the frequencies of 98~ and 102~, hence the breadth of the resonance curve, or, as it is also called, its BAND WIDTH, is 4~. The corresponding figures on Fig. 13C are 95~ and 105~, with a band width of 10~. (Again no attempt is made to draw illustrations in proper mathematical dimensions.) Greater damping produces greater band width and is perceived, when sufficiently great, as a 'blunting' of the sound. The opposite effect, reduction of blunting, is called 'sharpening'. Sharpness rather than bluntness is one of the attributes of a good, trained singing voice. A singer, then, must learn to eliminate as far as possible the damping effect contributed by the supra-glottal cavities and obstacles.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

IX

FILTERING

We may set up a battery of tuning forks, which we shall employ as resonators, and connect each with some kind of meter, as shown in *Fig. 14.* On my drawing the tuning forks are set

Fig. 14. Filters.

at intervals of 50∼. One may, of course, choose smaller or larger intervals, and place as many such tuning forks over as broad a total range as one wishes. Let us further see to it that each meter is connected with a stylus which, when activated, leaves on a piece of paper a visible trace indicating the approximate duration of its excitation. (I say 'approximate', because the storing of energy in the resonator will prolong its excitation beyond the activity of the sonator, but at a dwindling rate and with small power.) To obtain the element of time I shall, as I have done before (see Fig. 4), move the paper from below the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

styli at a regular speed from right to left. If a complex tone is now sounded, each tuning fork whose frequency is contained in the sound, be it that of a fundamental or of a sufficiently strong overtone, becomes a resonator, that is, it receives and perpetuates an impulse. If I play a note containing the frequencies 100~, 200~, 350~, and 450~, the meters of the corresponding tuning forks will react approximately for the duration of their stimulation and with a vigor commensurate with the power they received. The length of the impulse will be shown by the length of the trace each stylus leaves on the moving paper. The amount of power can, if one wishes, be translated into pressure of the stylus upon the paper in such a way that greater power would leave darker lines. However, while the frequencies can be read off along the scale, and the duration can be measured in inches and translated into seconds, it will be difficult to obtain an exact quantification of power from the degree of darkness of the line. A reading could be obtained by measuring the degree of darkness with some photo-sensitive device (comparable to the photographer's lightmeter), but this would require a complex process, and possibly would still not give precise acoustically translatable results. But this deficiency need not worry us in the analysis of speech, where, as I noted, the absolute amount of power is irrelevant.

Each of the resonators in this series of tuning forks, then, lifts out of the complex sound its own frequency. When a resonator is employed for this purpose it functions as a FILTER.

Now since the filters I have used are (ideal) tuning forks, every one of them is capable of filtering out only one frequency, namely, the one which it is capable of producing. Any other frequencies are simply not present as far as this particular set of filters is concerned. If, for example, I wished to filter and register the tone analyzed above (100~, 200~, 350~, and 450~) on a bank of tuning fork filters arranged at intervals of 100~ from zero, I should obtain registration only for the 100~ and 200~ components of our tone, while the 350~ and 450~ components pass unnoticed. Obviously the resultant

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

graph is in no wise a true analysis of the tone. Of course, if I knew in advance what frequencies the tone I wish to analyze contained, I could see to it that properly spaced filters of proper frequencies were used. But this is really putting the cart before the horse: after all, the purpose of an analysis is to find out what are the component frequencies of a complex tone.

The use of filters of the ideal tuning fork type has another disadvantage, which becomes serious when we analyze a damped wave, that is, one which is not produced by an ideal tuning fork or an instrument of similar quality. Speech sounds in particular are wholly unlike that of the tuning fork, for in speech sounds, which are damped sounds, not only the fundamental and its multiples make up the sound's quality, but also the frequencies which are contained within the band width of each component frequency. It is clear that any filtering which ignores these contributions to the composition of a speech sound and delivers a visual registration in which they leave no trace whatever, is unsatisfactory and may indeed be worthless.

It will therefore be necessary to equip a machine which registers speech by means of a filtering device analogous to that in Fig. 14, with resonator-filters which are sensitive to damping, that is, with such resonators as would themselves produce damped sounds if they were used as sonators. A filter of this kind is activated not only by the peak frequency of each damped component but also by other neighboring frequencies contained within the band width of its spectrum. One could, of course, take care of this requirement by employing a sufficiently large number of filters, possibly of the undamped variety and 1~ distant from one another. But this would necessitate a machine of extremely cumbersome proportions which might furnish huge graphs. There are simpler ways of obviating this difficulty, of which I shall speak later.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

X

ARTICULATION

We have by now gathered a considerable body of information on acoustic theory. The phenomena were discussed not in order to provide a theory of acoustics, but rather to demonstrate what role they play in the spectrographic analysis of speech: oscillography in general, resonance, reinforcement, damping, and filtering. We are now ready to apply this knowledge to spectrography.

After some general remarks I shall restrict my discussion, for the most part, to vowels, which are the most resonant, as it were, among resonant speech sounds. The other resonants, and the non-resonants, are equally susceptible of spectrographic analysis. But in their production moving organs other than the vocal lips, and air spaces other than those of the resonance cavities, participate; and the non-resonants contain significant elements of noise in addition to tones, or, in the case of unvoiced fricatives and plosives, noise only. Hence the spectrograms of the non-resonants will be based on oscillograms whose waves are non-periodic and non-repetitive, which I have not discussed. It is of course important for a spectrographer to acquaint himself with the acoustic and perceptual and spectrographic behavior of all these sounds also. But since it is my aim merely to introduce the reader to the principles of spectrography, I shall choose my examples and illustrations chiefly from the simplest as well as the most satisfying types of spectrographically analyzable sounds. Toward the end of this book I shall explain in a few words how the non-vowel speech sounds are registered on the spectrogram.

When in the process of expiration the air is pressed from the lungs through the glottis one may simply let it escape without

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

impeding its passage, keeping the vocal lips lax and open, and opposing no obstacles in the supra-glottal regions; that is, one simply breathes. Whatever noises one makes in the process are due to inevitable surface friction along the trachea, the mouth, and the nose, but they are linguistically meaningless or remain altogether imperceptible—although hard, deep breathing, approaching a snort or a sigh, may convey some meaning: relief, fatigue, exasperation, anger, excitement. But these are not articulate linguistic sounds in the terms of the definition of language, and may be ignored in the present context. During sleep these friction noises, especially those produced by the vibrating velum (soft palate), may escape control and become greatly magnified; then one snores.

But there are certain speech sounds in various languages which acoustically are no more than such noises caused by simple friction, or by explosion and subsequent homorganic friction, though by their linguistic function they are elevated from inarticulate noises to phonemic status. They are called fricatives and affricates, voiceless if the vocal lips remain inactive, like [s, f, ʃ, θ; ts, kχ], or voiced, like [z, v, ʒ, ð]. The so-called uvular *r* of Parisian French, northern German, northern Italian dialects, and Puerto Rican Spanish, for example, is in some of its forms, when not actually trilled, acoustically not at all unlike the noise that accompanies the clearing of the throat in preparation for spitting. And some languages have clicking and hissing phonemes made of noises which in our culture, if in recurrent use at all, may serve as inarticulate signals to tamed animals. (Note that I refer to them as signals. Only as parts of a linguistic system do they function as symbols.)

In terms of articulation with or without participation of the vocal lips one distinguishes between voiced and voiceless sounds. But for a spectrographically more suitable systematic description of what happens to the stream of air on its passage through the vocal apparatus it is better to classify phonemes in the following three classes which I have already mentioned: resonants, fricatives, and plosives (stops, explosives), the last

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

two being subdivided into voiced and voiceless subclasses. (All resonants are, by definition, voiced; voiceless and resonant are mutually exclusive qualities.)

Resonants. All resonants contain the glottal tone, though highly modified by cavity resonances. Vowels, laterals, nasals, and trills belong in this category. The obstacles put into the stream of air are minimal (vowels), or relatively weak (non-vowel resonants). For vowels, they consist only of the inevitable friction along the surfaces of throat and mouth which the moving air touches; laterals (/l/) have a continued but only partial closure in the center of the mouth, with air escaping along either side, or both sides, of the tongue; in the case of nasals the air stream is diverted, partly or fully, through the nose and must move through constricted and sinuous passages; in trills (/r/) there occurs a series of intermittent interruptions, full or partial, in the air stream. The effect of resonance is greatest in the vowels; they are, one might say, pure glottal voice, least affected by additional non-glottal frequencies, though most strongly modified by supra-glottal resonances.

The so-called 'voiceless vowels' (a contradiction in terms under my present definition of the vowel) of whispering and in some languages of normal speech (Japanese), which, as I said before, may be so termed because of their phonemic function but not in view of their phonetic nature, are really not tones modified in the supra-glottal cavities. They are, instead, frictional noises distinguished from one another by the same variations in size and shape of the supra-glottal cavities which would modify the glottal tone if it were present. In other words, a given 'voiceless vowel' has all the articulatory aspects of the normal vowel, except voice.

The non-vowel members of the class of resonants differ from the vowels in both the modifications of the glottal tone and the severally typical noise patterns caused by friction and obstacles along the passages. In the spectrogram, as we shall see, these noises appear as random frequencies, though of course each of

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the diverse phonemes exhibits its own typical randomness.

The following English phonemes are resonants: all vowels, /m, n, ŋ, l, r, j, w/.

Fricatives. In these sounds, the noises that may have occurred as accidental, unavoidable frictions in the resonants, are purposely emphasized and selected so as to become significant features by which the various fricatives are typically identified. That is to say, the place and type of a friction-producing partial closure or stricture are intentionally and typically chosen by the speaker, and may happen anywhere between the glottis and the lips. Consequently, the frictional features become acoustically significant and spectrographically prevalent. The articulation of a voiced fricative is accompanied by the glottal tone, albeit one not significantly modified in its overtones; it provides merely a background of voice. Voiceless fricatives are pure noises in which all features of the glottal tone are acoustically and spectrographically lacking.

The following English phonemes are fricatives: voiced—/v, z, ʒ, ð/, voiceless—/f, s, ʃ, θ/. (Some linguists prefer to consider also /w/ and /j/ as voiced fricatives. They partake of both classes, resonant and fricative, I think, and indeed the possibility of calling them semi-vowels or semi-consonants depending on their position after or before a vowel, reflects their dual phonetic character. I list them under resonants, because on the spectrogram they do show the features characteristic of that class.) I should add here also the voiceless spirant /h/, which one may call a fricative with minimal friction; if friction were totally absent, only quiet breathing would remain, which is not a phoneme.

Plosives (stops, explosives). These are produced, theoretically and ideally, by a complete closure which the speaker makes intentionally and typically somewhere between the glottis and the lips, and by means of which he completely interrupts the stream of air issuing from the mouth or nose. This is, then, a particularly severe case of stricture, and in a sense a plosive is a maximum case of a fricative. The popping

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

open of compressed lips, the sudden release of the tip of the tongue pressed against the teeth or gums, the loosening of the back of the tongue pressed against the hard or soft palate, the abrupt separating of the vocal lips—all these are articulations whose acoustic effects are used as speech sounds. They are, in order, labial, dental or alveolar, palatal or velar, and glottal stops. Preceding the release of the closure there occurs a moment of suspended articulation during which air pressure accumulates behind the obstacle. Paradoxically, then, plosives are sounds characterized by the absence of sound. But their onset and termination modify typically and perceptibly, according to the place of the explosion, the preceding and following sounds. The spectrogram will bear this out. On it, the registration of a stop is no more than a typically modified graph of the end of the preceding or of the beginning of the following sound, while the pre-explosion silence shows up as a blank space, as the complete absence of spectrographic activity, at least in the voiceless stops. The voiced stops, however, are accompanied throughout their articulation by a background humming of voice, due to the vibrations of the vocal lips. But all other articulatory activity remains suspended. On the spectrogram this voicing is discernible, but only as a secondary, insignificant feature which, while attesting the articulation of a voiced sound, does not help in identifying the kind of sound or the type of articulation.

The following English phonemes are stops: voiced—/b, d, g/, voiceless /p, t, k/.

The fundamental frequency of the glottal tone, wherever one is present in the stream of speech (that is, in all voiced sounds), provides of course the natural pitch of the normal speaking voice of the individual. It can be modified at will in speech and especially in singing by varying the tautness of the vocal lips, which are in effect tuned by the speaker or singer as are the strings on a violin by the violinist. Stopping, which varies the length of the string and is essential for performance on all stringed instruments of the violin type, is of minimal

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

importance for the vocal lips. It consists, not in real stopping by some organ which, like the finger on the string, may shorten the vibrating portion of the vocal lips, but merely in allowing only a part of them to vibrate. While no theory concerning the peculiar performance of falsetto singing or speaking and yodeling has found common acceptance, one hypothesis suggests that in this kind of sound production only the center portions of the vocal lips vibrate, producing a tone, generally a bit 'thin', that lies above the normal pitch range of the singer's or speaker's voice.

Now if voice pitch were a significant factor in linguistic communication on the phonemic level, only persons of equal natural pitch could communicate with one another, or else all speakers of the same language would be constrained to make equal communications at equal pitch levels, that is, speaking would always be, in a way, a kind of singing. Neither of these conditions prevails, the first being biologically impossible, the second socially too demanding. I have already mentioned that even in languages where pitch is significant, where the meaning of a word can be changed by variations in pitch while otherwise the phonemes remain the same, as in Chinese, only a relative variation of the pitch contour or the pitch level, and not adherence to an absolutely fixed musical melody, is required. The same is true where intonation provides meaningful signals, for example, by distinguishing questions from affirmations: *He says so. He says so?*

Instead of following a linguistically predetermined melodic line at a prescribed pitch, speakers actually utter all vowels, for example, at whatever pitch they choose or is most convenient, with each of their articulations being linguistically identified by features other than glottal pitch. But since we have learned that vowels are the products of the glottal tone, each typically modified in its overtones but not otherwise noticeably or significantly varied, it follows that the linguistic differences in the vowels which every speaker is capable of producing and obliged to employ, must be due to the articu-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

latory and acoustic action of the supra-glottal modifications alone, that is, to resonance and reinforcement in the supra-glottal cavities. In other words, /a, e, i, o, u/ etc. owe their distinctiveness not to what we do with the vocal lips but to what we do with our mouth.

And this brings us to another requirement for the machine we wish to construct. It must be able to register speech in such a way that absolute glottal pitch is properly relegated to its position of irrelevance for linguistic communication, if not eliminated altogether. The machine must, rather, pay particular attention to the supra-glottal modifications and events. It is clear that a mere oscillographic representation of speech sounds does not exclude glottal pitch from its registration: in the oscillogram of a complex sound, pitch is amalgamated with the other components. Attempts have been made to separate the glottal and the supra-glottal oscillations by representing every one of them in a separate band. But since this is generally accomplished by the crude and disturbing expedient, as required by the kymograph, for example, of attaching several devices to several parts of the speaker's speech organs for simultaneous but independent registration (and a number of speech organs are not accessible at all in this way, at least not without distorting or impeding their normal functioning), a true and mathematically exact separation is impossible. However, if a sound is analyzed in the spectrographic way, then each frequency is registered separately, together with some indication of its power. Now the glottal pitch can be easily identified (either as the lowest frequency present or, in its absence, as the distance between the harmonics), and ignored, whereas the other frequencies and their distribution can be given their proper weight either singly or, as we shall see, more significantly in groups.

It is true that this separation of frequencies is of importance only where there is a glottal tone present, that is, in the resonants and in the voiced fricatives and plosives. But the spectrographic identification of the sounds without glottal participation owes much to the configuration of those with it, and, vice versa,

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the shape of the voiced sounds is in some measure determined by the surrounding voiceless ones.

We have up to now established three desiderata in which the spectrogram will have to render, in comparison to an oscillogram, a truer visualization of speech for linguistic purposes. (1) It must omit the non-relevant feature of phase; (2) it must separate the fused frequencies from one another in such a way as to let the experimenter distinguish the significant from the non-significant components; (3) it must accomplish this in a manner similar to the process of hearing itself.

Point (3) requires some comment. The movable parts of the ear—the eardrum, the tiny ossicles of the middle ear (hammer, anvil, and stirrup, so called after their shape), and the fluid contained in the cochlea of the inner ear—actually pass on the vibrations of the air as a physical movement. Somehow, through a process not yet wholly understood, but in any event through the intervention of the nerves and the brain, we perceive ultimately not a vibration of air particles but a sound. (It is, of course, true that in terms of physics vibration *is* sound. But it is equally true that we do not perceive the vibrations as vibrations, that is, as a sequence of condensations and rarefactions of the air, but as something wholly different). One part of this operation is surely a process of filtering and harmonic analysis which extracts and transmits separately the various frequencies which impinge upon the ear as a unified complex sound. Possibly the hair cells and the nerve ends of the inner ear, which protrude into the cochlear fluid where the vibrations have been ultimately conveyed all the way from the eardrum, operate as a set of resonator-filters, something like the battery of tuning forks of Fig. 14. The filters of the ear are very numerous and are tuned to a very great variety of frequencies. Even if each of them were an ideal filter that passes only one frequency and not a whole band of them, the ear, thanks to their number, would still be able to catch an enormous variety of different frequencies which would lie so close together that they could cover the whole spectrum within the limits of

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

human sensitivity. Whatever the operation in detail, filtering and analysis take place, it seems, in such a way that the brain is notified of the presence of certain amounts of power at certain frequencies. This is, mutatis mutandis, also the manner in which the spectrograph operates.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XI

PHONEMIZATION

Here I must point out a certain disability from which the spectrograph we are constructing, as well as every other machine for language registration, must unavoidably suffer, for the very good reason that, although they are 'educated' machines and are provided with a substitute nervous apparatus and a substitute brain, they lack cultural attitudes and adjustability; they can make no judgments and evaluations beside those with which their creator has been willing and able to provide them. To elucidate I must review briefly some phonemic facts.

The configuration of cavities and vocal organs responsible for the variety of speech sounds is, in addition to being determined by volitional choices and variations, also dependent on the individual's anatomic constitution, such as the shape and size of the head, mouth, nose, and the various organs in them. Yet each person can neutralize his personal peculiarities to such a degree that, while no two speakers ever produce *acoustically* same sounds, while not even one person can pronounce in succession such same sounds, everyone can learn to make *linguistically* 'same' sounds. This kind of sameness is attained when articulations fall severally into classes of sames, which are called phonemes. The physical limits of such phonemic sameness are defined by the structure of the language, by the specific way in which any given language classifies the continuum of phonetic reality into significant, distinct segments, and by acoustic and physical attributes.

This segmentation of the acoustical continuum of frequencies of sound may be compared with the segmentation of the optical continuum of frequencies of light, such as seen, for example, in the rainbow. In our western cultures we conventionally choose

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

to cut the continuous solar spectrum into seven visible so-called colors: red, orange, yellow, green, blue, indigo, and violet. Note that on a real spectrum, such as that produced by light passing through a prism, these colors shade into one another, and that within the segment we have agreed to call 'orange', as distinguished from a piece of paper dyed in a uniform 'orange', the color is not uniform throughout, and that the areas bordering on the neighboring segments on either side, red and yellow, are actually 'red' and 'yellow' as much as they are 'orange'. There is really, then, no physical reason why the boundaries could not be slightly moved, or why indeed one could not divide the continuum into differently bounded or more or fewer segments—and certain societies and cultures do just that. But whatever the differences of segmentation that arise from these divergent habits, they do of course not reflect or condition the physical ability of the viewer. They may, however, cause cultural prejudices and a cultural disability, so that members of two different 'color cultures', as it were, do not agree on the place and number of the boundaries which cut the spectral continuum, hence do not agree on what 'same' colors are. While there can be no question, therefore, of general physical colorblindness among all members of a culture, there may occur what one could call a cultural colorblindness in terms of the scale of another color culture.

Similarly, the conventional segmentation of the acoustic continuous spectrum of frequencies into the speech sounds of a language, while it cannot cause a real physical inability of a speaker to perceive and produce all the sounds which a speaker of another language pronounces, may result in the speaker's cultural inability to perform as hearer or speaker commensurately with his physical capacity. This may lead to a kind of cultural (phonemic) deaf-muteness so strongly ingrained that an individual may have extreme difficulties in acquiring foreign speech habits or, in fact, never attains full, that is, native mastery of them. He may even abandon all efforts to improve his speech if his 'foreign accent' does not interfere with his

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

making himself understood and if he has no trouble understanding the native speakers, if, in other words, incomplete adherence to the acquired cultural pattern is not penalized, or at least penalized no more than the speaker is willing and able to bear. (The limits of tolerance are in such cases determined by the personality involved, by the culture in which he moves, and by the purposes to which he puts his language. The United States for example, and in particular a metropolis like New York, is on the whole tolerant of foreign accents. Still, a French actor, no matter how excellent an interpreter of Corneille he may be, would be unbearable as King Lear if he spoke his lines with a French accent, and he would be penalized by the public surely beyond the limits of his endurance.)

Sometimes linguistic sameness seems liberal and generous in comparison with acoustic sameness. For example, sounds which in Indo-European languages we so fully and easily classify as different phonemes, like /r/ and /l/, are in Japanese phonemically indistinguishable variants, or allophones, of one phoneme (which accordingly could be transcribed as either /l/ or /r/, or, for that matter, by any written symbol agreed upon); or, rather, Japanese has a phoneme which acoustically and articulatorily is so located that the sounds we call [l] and [r] in English are so near it that they both appear to qualify as realizations thereof. This does not mean, by the way, that Japanese 'confuses' /l/ and /r/: it is nonsense to say that one language confuses the phonemes of another language; it means that Japanese has no /l/ or /r/, and that one phoneme it does have, call it /x/, is acoustically and articulatorily related to both /l/ and /r/. This is possible because English /l/ and /r/ are, to begin with, phonetically not very distant from one another. To English speakers they are sufficiently distant to be distinct, to Japanese speakers they are not. But, in the sense of what I said concerning phonetic similarity of all realizations and allophones of a phoneme, I do not think that any language should be so analyzed as to put two sounds that are acoustically as distant as, say, [h] and [ɬ], or [g] and [s] into one phoneme.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Of course, allophones are not always freely variable. They may occur as positional variants in such a way that two allophones of a phoneme are used predictably in two different positions, as the two allophones of /l/ in the word *lull* already mentioned. A foreigner who brings from his native phonemic inventory one single /l/ phoneme (unlike the Japanese who does not have anything that matches an English /l/ phoneme, even though a linguist may choose the transcription /l/ for what the Japanese does have) cannot err phonemically in his pronunciation of English. That is to say, no matter what kind of /l/ he uses, he will pronounce the word *lull* intelligibly and unmistakably, even without the aid of a meaningful context. But if his language has no positional variants of /l/, as is true of French or Italian, or if his variants of /l/ do not match the kind of complementary distribution they have in English, as is true in German, the foreigner may speak with what is called a foreign accent. This means, then, that although he may still be understandable (though I am sure that wrong allophones may at least impair if not prevent intelligibility) his pronunciation will sound strange because he does not use suitable variants, or employs them inappropriately.

All this points to the fact that a phonemic transcription is really worthless, unless it is understood what phonetic values attach to the symbols that we use. And if we use, by and large, symbols borrowed from the International Phonetic Alphabet, a certain phonetic value is implied in each; otherwise we might as well use numbers, or any kind of fancy symbols like / ξ ⊥ □/—and sometimes this would be a salutary thing, in particular if it helps to make clear that English /l/ and Japanese /l/ are not sames. In any event, the presence of acoustic values in every symbol we use in our phonemic transcriptions, though they may difffer from language to language, implies that, whether we admit it or not, some phonetic reality is present in our phonemic analysis. And if this is so, it will be easy to concede the merit of assessing phonetic similarity in phonemic classification.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Now it may occur that a speaker fails to distinguish in his pronunciation of a foreign language not just the allophones but the phonemes, like the Spaniard who knows only one /i/-type of phoneme and therefore in his speech does not automatically separate English /i/ as in *beat* from /I/ as in *bit,* and who may indeed completely fail to make himself understood. Similarly, knowing only one /e/-type of phoneme, he may not properly distinguish in his speech *bate, bet,* and *bat.* The inevitablity of a breakdown in communication is often precluded, or its severity happily alleviated, since the hearer is able to perform what communication engineers call code switching, a process of adjustment to the articulatory habits of the speaker which permits the listener to learn quickly certain types and degrees of phonemic deviation for-which he must make allowance. And, of course, context is of immeasurable help in restoring the proper intelligibility of an utterance which, on acoustic grounds alone, one would expect to remain incomprehensible. But reliance on adroit code switching and contextual aid is neither practically safe nor theoretically sound, at least not in analysis on the phonemic level.

Now a machine has no language, hence it cannot make by itself the distinction between acoustic and linguistic, between physical and cultural, sameness. While it cannot suffer from cultural sound deafness and therefore, fortunately, cannot distort or misunderstand the physical data it receives, it can also not order them or analyze them in accordance with any linguistic conventions. We do not construct and operate an English spectrograph, or a French kymograph, or a German cathode ray oscilloscope any more than we produce Fahrenheit or centigrade mercury for two kinds of thermometers. Just as in all thermometers the mercury behaves physically exactly alike, and we ourselves add a conventional man-made scale of one type or another in accordance with the culture in which we live, so the sound registration machines and all their functioning parts behave exactly alike no matter what is fed into them, and we merely add our interpretation, our scale, to the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

registered data. A spectrogram conveys, therefore, acoustic or, if transferred to the linguistic plane, phonetic information, but not phonemic information. In order to derive the latter from it, it is necessary for the viewer or the experimenter to have established his phonemic scale beforehand, just as the reading of what we call degrees of temperature on a thermometer can be accomplished only after the mercury column has been provided with a scale of one kind or another. That is to say, the spectrographer must know the language, the code, which provides linguistic sense and classification for the acoustic data. The spectrograph has no part in determining what is, or what should be, a phoneme, except insofar as profound acoustic and spectrographic dissimilarity of two sounds might instill in the spectrographer a reluctance to class them as one phoneme. But once the scale, of phonemes or degrees, is established and calibrated in the machine, then the spectrograph, or the thermometer, may be employed for applying this scale to all events to which it pertains, that is, to measure and classify all events and physical data which the machine perceives. In other words, the machine can be instructed to pay attention to whatever data we want it to observe and isolate; but in order to have the machine behave according to our wishes, we must first know ourselves what we are looking for. It is obvious that to the machine, be it a spectrograph or a thermometer, the registration that it performs makes no sense and has no purpose. A spectrograph knows no more about phonemes than does a thermometer about the boiling point of water at sea level. Meaningfulness is imputed by the user. It is imputed, in the case of the spectrograph analyzing speech, by the physicist and the acoustician only in his capacity as a linguist—from which it follows that a physicist or an acoustician must be acquainted with certain indispensable aspects of linguistics if he wants to be a spectrographer of *language*.

By the same token, of course, the linguist who wants to talk about spectrography and, to be sure, about phonetics and phonemics, must know the rudiments of acoustics and spectro-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

graphy, at least to the extent that they are discussed in this book. If he does, he will be less inclined in his phonemic analysis and schemes to apply criteria and procedures of armchair phonemics, which has a way of flouting or ignoring the physical reality of language; he will become reluctant to put all his eggs into the relational, distributional basket, an imprudence I noted before. Such an attitude of realism and caution will, in my opinion and that of many others, benefit linguistic science.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XII

VOWEL PHONATION

I shall now choose three English vowel phonemes, /i, a, u/, and demonstrate their articulations in some detail so as to derive from them a clear view of the size and shape of the oral cavities which I have mentioned so many times. (It is customary to speak of the articulation, analysis, etc. of phonemes. Strictly speaking, this is a lax use of terminology, since a phoneme is a class of things of which only the realizations, or allophones, can be articulated, analyzed, etc. But the usage is common, and I shall make no special effort to avoid it. Little harm is done as long as all statements are understood with this reservation.) *Fig. 15* represents a longitudinal section of the oral cavity. In the production of the vowels the vocal lips oscillate and send out the complex glottal tone, which includes the pitch of the voice. The position of the tongue can be ascertained through X-ray pictures; the position of the lips and lower jaw is visible from the outside. (An excellent view of the operation of the vocal organs is afforded by films of subjects who, through disease or accident, have lost the outer integuments of the vocal apparatus so that their larynx and supra-glottal cavities are open to undisturbed sight.)

When /a/ is pronounced the jaw is lowered, the tongue is flat at the bottom of the mouth (hence the doctor who wants to examine the patient's throat asks him to say "a", thus getting the clearest view with the least use of the unpleasant tongue depressor), the lips are wide open, the teeth farthest apart. The oral cavitiy is one single hollow, maximally large.

In the production of /i/ the jaw is raised, the tongue is humped forward toward the anterior (hard) palate, the lips almost touch and their corners are drawn sideways, the teeth are

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

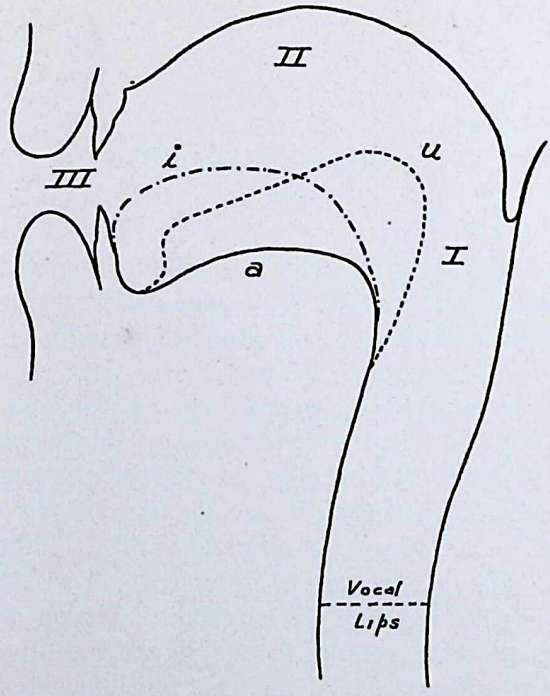


Fig. 15. Oral cavity.

nearly closed. The oral cavity is divided into two parts by the tongue, the smaller before and the larger behind the tongue, with a third small cavity lying between the teeth and the lips.

For the /u/ the jaw remains raised and the teeth nearly closed, but now the tongue humps backwards toward the posterior (soft) palate; the lips are rounded and puckered. The oral cavity is again bisected, but this time the larger portion is before and the smaller behind the tongue, while the third cavity between the teeth and the lips has been considerably enlarged, thus giving the entire mouth a lengthened, somewhat funnel-shaped profile.

Each of the three sub-cavities, which I marked I, II, and III, acts as a resonance chamber, the enclosed air spaces being the resonators whose acoustic properties are determined by size

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

and shape, each having, like the air enclosed in a bottle, a natural frequency at which it can be made to sound or resound. If any of the frequencies associated with the cavities is present in the tone that rises from the glottis, as it most likely is in view of the complexity of the glottal tone, the cavity or cavities concerned will resonate. This resonance will alter the frequency-power composition of the glottal tone, that is to say, it will change the quality of the glottal tone. (In addition, as I mentioned before, the glottal tone will be modified by the addition of some frequencies arising from friction and vibration of the fleshy and bony parts of the walls of the pharynx and the mouth, but these contributions are on the whole negligible in characterizing resonant sounds, especially vocalic speech sounds.) Also, some reinforcement will take place. It is obvious that the phonemes /a, i, u/, produced under such radically varied conditions of resonance and reinforcement of a highly complex glottal tone, will deliver, even at the same glottal pitch, oscillograms and spectrograms which differ from one another significantly and, as we are dealing with classes, typically.

Since vowels are, it appears, sufficiently and significantly (though not fully!) described in terms of articulation by the parameters of the position of the tongue, jaw, lips, and teeth, one might reasonably ask whether on a spectrogram a representation of the same four parameters would sufficiently and significantly illustrate the phonemic distinctiveness perceived by the speaker of a language which has these three phonemes.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XIII

SPECTROPHONES

In designing the spectrum of the tone shown in Fig. 11, I did three things: first, I selected a very simple wave (in comparison with speech sounds), made up of only two partials; second, I chose a sustained sound, with an infinitely repetitive periodic wave, so that, by selecting any one point of time in its course, I could obtain a true picture: this sound has always a certain power at 200~ and a certain power at 600~, hence the sketch of Fig. 11 is a complete and exact spectrographic rendition of it; and third, I employed an undamped sound.

But in human speech, none of these three conditions, which I accepted for the sake of simplicity in most of my illustrations so far, prevails. In analyzing speech and speech sounds, therefore, one is faced with greater intricacy on all three counts: first, power is present not at two but at a great number of frequencies (see the complexity of the glottal tone); second, the element of time is of the essence since no speech sound is normally sustained over more than a fraction of a second and since all speech sounds derive their typical spectrographic shape, as we shall see, from the time dimension; third, speech sounds are damped sounds and must be analyzed in such a way that the effect of damping becomes visible in the graph.

To perform an analysis that does justice to all these complexities leisurely and with proper accuracy, we cannot just ask a person to speak with convenient slowness and to draw out his phonemes so that we may obtain a clearer picture of at least some of them, despite their fleetingness. The reason is that deliberate extension of articulation in the mouth of a speaker alters the true, normal acoustic composition of a speech sound and distorts it. (There exists a way of stretching sounds, but

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

this is accomplished by a controlled device, called the speech stretcher, that operates from an original natural speech recording.)

The first two requirements I just mentioned can essentially be satisfied by the kind of arrangement illustrated in Fig. 14: a battery of resonator-filters tuned to a selected number of frequencies at chosen intervals; and a moving strip of paper on which the presence of power is registered in each filter as a function of time. We do thereby give up the precise quantification (but not necessarily the approximate indication) of a third parameter, the amount of power present, because we must obtain a two-dimensional rather than a three-dimensional visualization. The third requirement, concerning proper attention to damping, can be met by employing appropriate filters, that is, damped resonators. And a suitable and controlled articulation, capable of exact repetition of the sample of speech to be analyzed, can be obtained by using a recording in the place of a live speaker.

I shall now show a series of spectrographic illustrations of increasing complexity, which lead us step by step toward fulfilling all the conditions and requirements for a complete spectrogram mentioned so far, and which the spectrograph will have to meet.

Fig. 16 represents a spectrographic sketch in which each line on the horizontal marks the presence, and the duration, of power present at frequencies that are arranged on the vertical. Note, however, that I have chosen again, for the moment, a succession of sounds of which each has but one frequency, a fundamental, and no overtones. Obviously these are not sounds such as most instruments produce, less even speech sounds, but rather the tones of successively struck ideal tuning forks, or of an ocarina, or of human whistling. The left half of the illustration registers a melody of four tones played consecutively, or staccato. The right half is the so-called 'wolf-whistle', by which young men express vocal though inarticulate admiration of a pretty girl; hence it is not a staccato sequence but shows

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

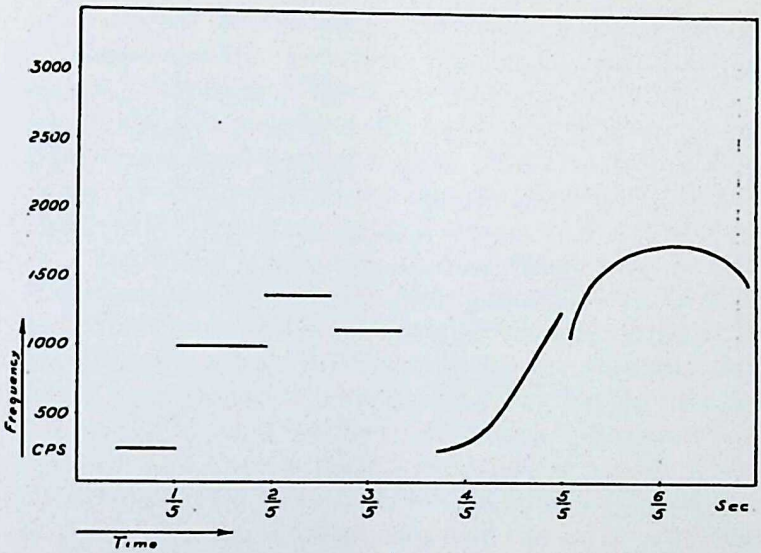


Fig. 16. Spectrographic sketch of simple tones.

sliding from one frequency to another. Speech will be mainly of the gliding type.

The next two illustrations are spectrograms of more complex tones. *Fig. 17* is a melody played by a violin, and *Fig. 18* is the wholf-whistle of Fig. 16 played on a sliding trombone. Not all overtones actually contained in the tone of either instrument are shown in the sketch. The violin tones are played staccato, the trombone plays a sliding legato. Both spectrograms are made by fixed ideal filters, such as allow but one frequency to pass. (The increase in steepness or relative height of the curves of the higher as compared with the lower harmonics, which accompanies a rise of the fundamental pitch, is due to the fact that overtones, being multiples of the fundamental, increase their frequency at a more rapid rate than does the fundamental. If we read the fundamental at two successive points in time as 10~ and 30~, the second harmonic at the equivalent points is at 20~ and 60~, the tenth at 100~ and 300~; this means that

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

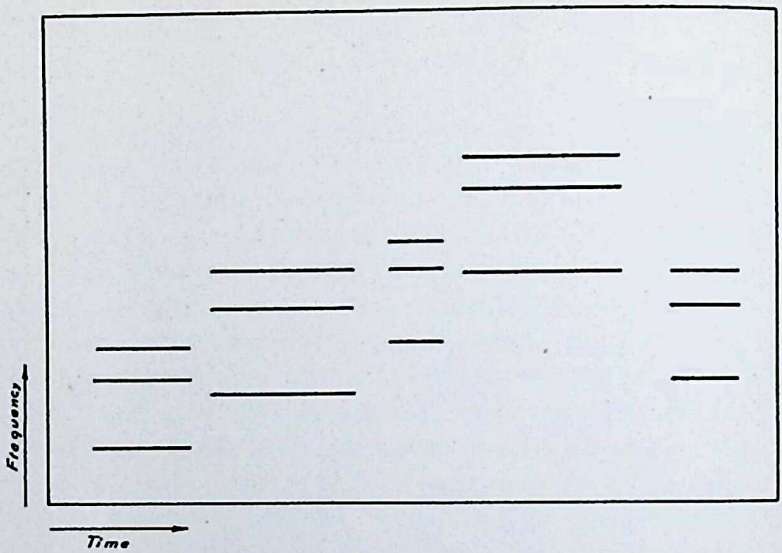


Fig. 17. Spectrographic sketch, complex tones, staccato.

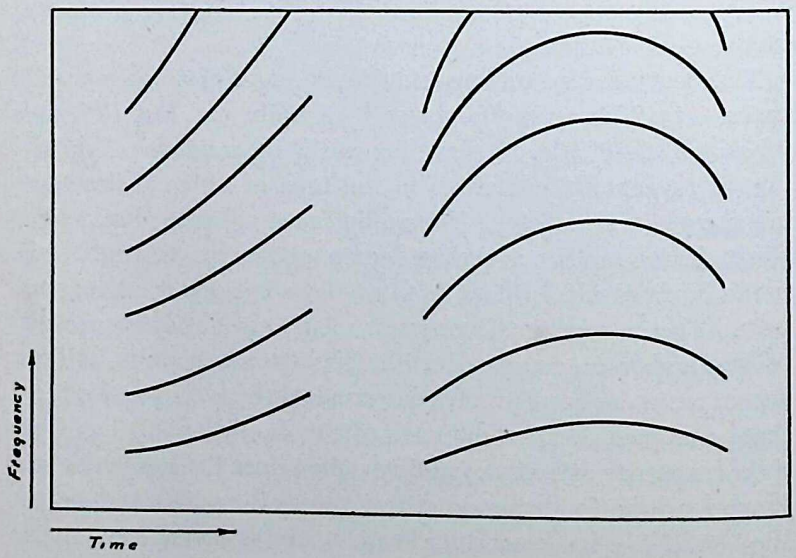


Fig. 18. Spectrographic sketch, complex tones, legato.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the difference between higher and lower frequency level at the fundamental was only 20~, grew to 40~ for the second harmonic, and reached 200~at the tenth.)

Since the paper, on which the registration of power present is performed by the activated stylus, moves horizontally from right to left, the stylus can leave on the paper only horizontal traces, whose length represents the duration of activation of the stylus, that is, the time throughout which power is present at a given level of frequency. This means that only the horizontal tracings on the spectrogram are actually produced by the activated stylus. The lines showing the trombone harmonics, however, as well as those representing the whistled tone of Fig. 16, are drawn by hand, connecting the momentary occurrences of power at the filtered frequencies, which at each filtered level could appear only as dots at the proper distance in cycles from the abscissa. Obviously the stylus can make markings only at the points where it is actually activated by energy passing through a filter, with the other non-filtered frequencies passing unnoticed and unregistered. And since we are dealing here with a legato up- or down-sliding performance, each filter is excited only very briefly at the moment when a rising or falling harmonic traverses its frequency. This manner of completing a curve by connecting spaced dots that constitute parts of it, is a legitimate process, provided that the gaps in time between the dots are narrow enough to make the course of the curve a not unreasonable guess. (For example, a fever curve made from a patient's temperature chart is valid as long as enough readings are taken during the time which the curve represents. It is scarcely legitimate to take a person's temperature twice in one day, at 6 A.M. and at 6 P.M., say, and to connect the two points by either a straight line or a fancied curve.) Even if one assumes, however, that the shape of the hand-drawn curves on Fig. 18 is wholly justifiable, it is obvious that an experimenting scientist had better not be satisfied with a machine whose operation leaves that much to the discretion of the observer. We can complete the spectrum of

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

a trombone's tone with a fair amount of certainty in the manner shown in Fig. 18, since we know that, due to the peculiarities of the instrument and the tune played, a simple up- or down-gliding takes place. Speech sounds, on the other hand, which are so much more complex and have frequencies changing so quickly and so significantly, are not susceptible of such filling-in operations by the observer because chances of error are incomparably greater. Hence a further requirement for our spectrograph is that it provide us with a continued graph which visualizes both legato and staccato changes of frequencies correctly and automatically.

On spectra such as those of Figs. 17 and 18 an approximate indication of power changes and differences could possibly be indicated by varying the blackness of the lines. This would represent the varying pressure of the stylus on the paper, or, in the case of a stylus that burns dark traces into a sensitized paper, the varying amounts of electric voltage by which the stylus is excitated :the greater the power present, the greater the pressure, or the voltage, and the darker the stylus trace.

In the spectrographic analysis of speech, the successive segments of the speaker's utterance which are called (allo-)phones of the phonemes of the language he speaks, will leave their several visual traces upon the graph paper, exactly as did the successive notes of the musical performances recorded on Figs. 16, 17, and 18. These optical representations of linguistic segments I shall therefore call SPECTROPHONES. Given the complexity and the variability of speech sounds it is to be expected that spectrophones will be considerably more complicated registrations than those discussed so far.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

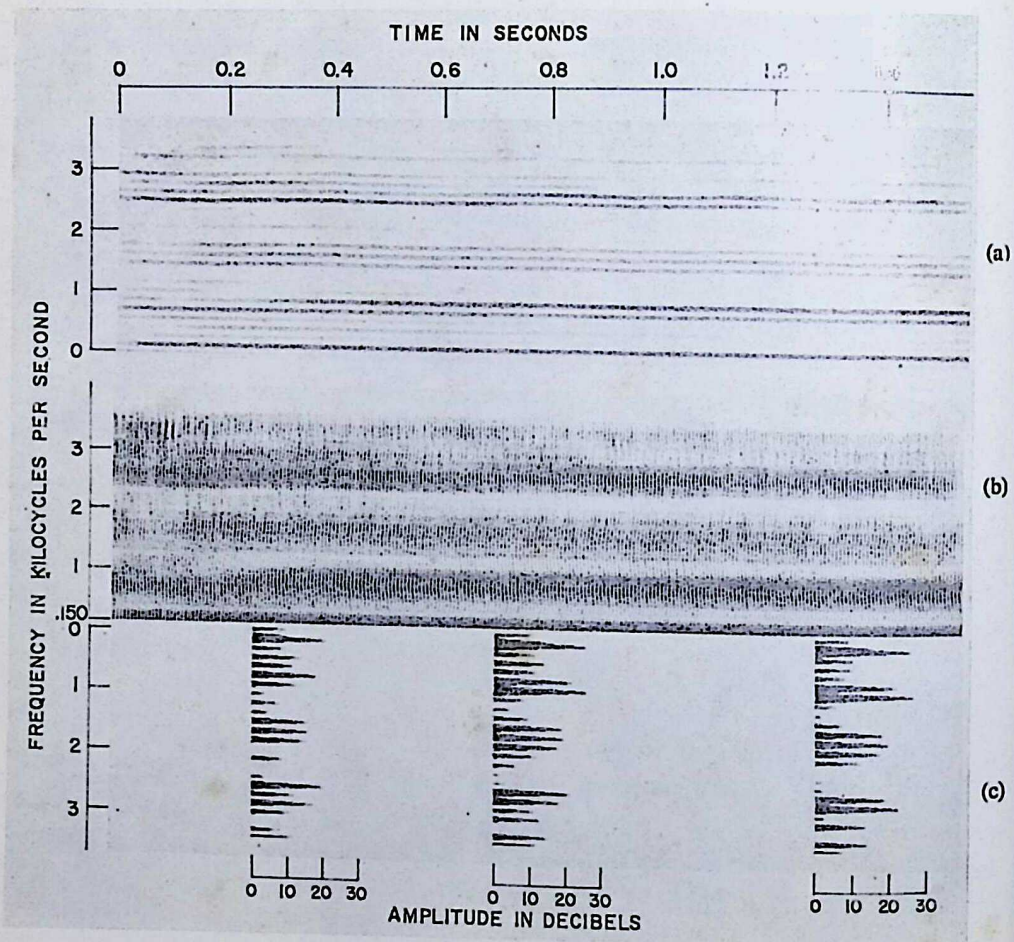


Fig. 19. Spectrograms of sustained cello tone (150~).
(a) Narrow band, 50~; (b) broad band, 300~; (c) amplitude sections.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XIV

SPECTROGRAPH

The upper and middle graphs of *Fig. 19* are actual spectrograms made on a spectrograph. How they are made and how they can be read I shall describe in detail in the subsequent chapters. For the moment I shall merely say that they visualize a sustained musical tone of 150~, performed by a cello, and show the frequencies of which the sound is composed between zero and 3600~. (1 kilocycle equals 1000 cycles.) Although overtones higher than 3600~ might be of interest in this instance, greater frequencies will not be necessary for the analysis of most speech sounds, except for experiments requiring additional precision. The filtering occurs at intervals of 18~. That is to say, on a total width of 3600~ there will be $\frac{3600}{18} =$ 200 registrations of frequencies on the ordinate. The time is registered on the abscissa; it is here about 1.5 seconds; we can record as much as 2.3 seconds on each single spectrogram, the maximum length being determined by the technical requirements of the machine.

The graph on top of Fig. 19 is a 50~, or narrow band spectrogram, so called because the filtering device which is employed reacts not only to one frequency but to all frequencies lying within a band 50~ wide, gathering power present at 25~ below and 25~ above its central frequency. For such a purpose we must use, it will be recalled, a filter whose physical properties correspond to those of a damped tuning fork, examples of which are shown in Fig. 13B and C, and not to those of an ideal tuning fork, as in Fig. 13A. Apart from band width, this spectrogram corresponds in principle to those of Figs. 17 and 18. The gray lines on it, representing the funda-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

mental (at 150~) and overtones present (at multiples of 150~), as distinct from those on Figs. 17 and 18, are unbroken and horizontal because the cello tone spectrographed is legato and sustained at the same pitch.

The middle graph is a spectrogram of the same tone made with a filter 300~ wide. The reason for its appearance, and its purpose, are discussed in detail in Chapter XVI.

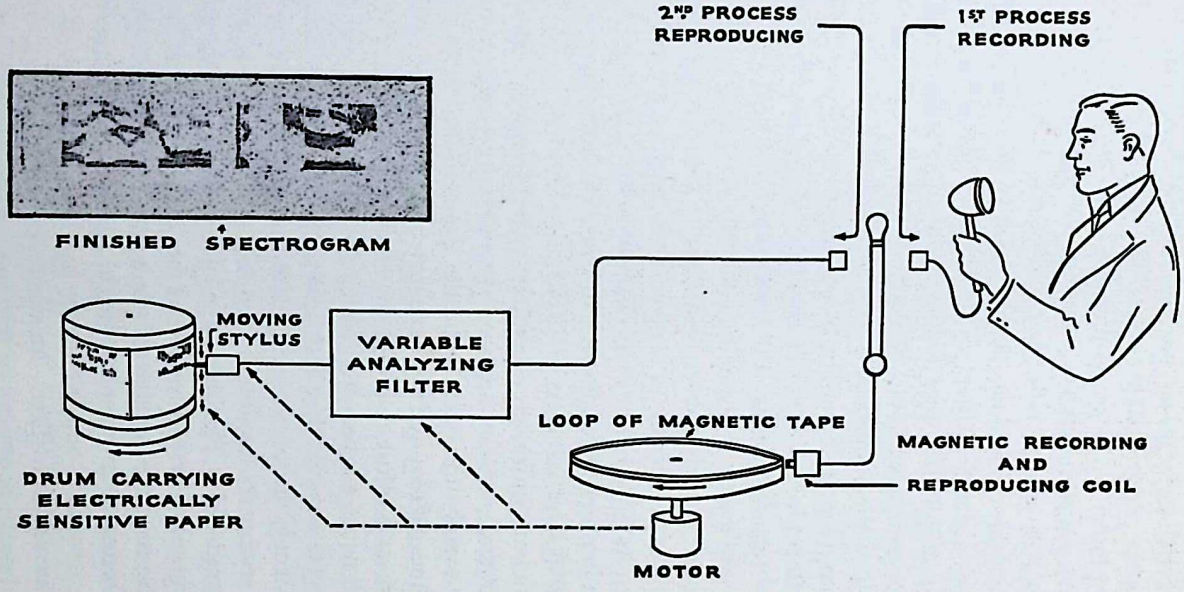
The bottom graph of Fig. 19 shows three so-called amplitude sections. They are made by the same machine (after certain readjustments), and represent what is shown in Fig. 11, the spectrum of the sound under examination: the place of the time dimension is now taken by the power dimension. Consequently, each of the three amplitude sections is a snapshot of the tone, taken at the moment of the time scale occupied by the vertical baseline of the section. Since the cello tone is sustained, the three snapshots are indentical—at least to the extent that the cello player could produce an unchanged, unwavering tone during $1\frac{1}{2}$ seconds. Amplitude sections of speech, in which neither frequencies nor powers are sustained, will not show this feature. In fact, every fraction of a second changes the spectrum; hence in spectrography of speech, if one adds an accompanying graph of amplitude sections at all, only a few selected positions can be shown.

Fig. 20 is a schematic of the spectrograph, the machine which produced the spectrograms of Fig. 19. It will not be necessary for the present purposes to enter upon the technical details of its construction and operation; but I shall discuss whatever the linguist should know.

The piece of paper on which the spectrogram appears is a little over two inches wide, with the actual spectrographic representation covering two inches exactly. The paper is covered with an electrically sensitized film into which the registering stylus burns a dark trace. (If the paper is carelessly handled the marks will smear.) It is wrapped around a cylinder which rotates in synchronization with a tape loop on which the recorded message is placed. This record is played repeatedly

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri



THE SOUND SPECTROGRAPH

Fig. 20.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

(two-hundred times) during the analysis. The reason for this particular arrangement is the following.

If we employed in the spectrograph which made Fig. 19, so crude a device as the battery of tuning forks of Fig. 14 (albeit damped, that is, of a certain filter width), the machine would have to contain the unmanageable number of two-hundred separate forks or other filters. This engineering problem was solved by equipping the apparatus with a single filter that has a modulator and a variable oscillator; that is to say a single resonator-filter that adjusts itself, in synchronization with the machine in operation, to various frequencies successively, filtering one after the other. This requires, of course, that the message to be analyzed by the spectrograph is fed into the machine not just once for simultaneous filtering at all frequencies, but once for every successive filtering at one frequency at a time, with each filtering at a given and constant interval (in cycles) from its predecessor and successor. Our spectrograph is set for two-hundred filtering runs to cover the total frequency width of 3600~, at intervals of 18~.

Now a human informant cannot be relied upon to repeat a speech utterance, or any sung or played tone, twice in succession in exactly the same acoustic manner, let alone two-hundred times as required by the spectrograph, to say nothing of the time and effort required for such a performance. To provide the necessary acoustic input to a spectrograph capable of analyzing a sample of speech lasting 2.3 seconds, a human performer would have to repeat his utterance, without in the slightest varying it, two-hundred times in about $7\frac{1}{2}$ minutes, which is impossible. Therefore, the piece of speech that is to be spectrographed is recorded on tape, and this recording is then fed the required number of times into the analyzing part of the spectrograph, where at each successive playing one frequency, or rather one frequency band, is lifted out of the total by the variable filter. At whatever point in the whole utterance the one or several of the frequencies passable by the filter at a given moment are present, the movable stylus whose

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

course along the width of the drum is synchronized with the progress of the filter, is activated and records the impulse on the paper.

Both the loop and the drum perform, then, two-hundred revolutions for the analysis of each recording. There is nothing peremptory about this number; it merely was found to be convenient and satisfactory. A larger or smaller number of scanning runs would be possible. Also, the best duration for the piece to be analyzed in one operation has been found empirically to be 2.3 seconds; it is determined by the requirements for a reasonable size of the machine and a reasonable length of time for the analysis. The total time necessary for the completion of a spectrogram would therefore be about $7\frac{1}{2}$ minutes; but the process can be speeded up, without causing distortion in the graph, to the point where it takes but $2\frac{1}{2}$ minutes.

With each revolution of the loop and the drum, that is, with each repetition of the message, the stylus completes one circuit around the drum. However, this is not a true geometric circle; for the stylus does not run along the same level and does not then jump the interval of 18~ at the end of each trip around, but rather follows a gradual course covering in each run the 18~ interval, so that at the end of one revolution of the drum it will be 18~ higher than at the beginning, in proper position for the start of the next. In other words, the stylus actually describes a spiral. If, therefore, the stylus starts at zero, it ends its first run at 18~, then on the next run moves from 18~ to 36~, and so on, with the two-hundredth run covering the area between 3582~ and 3600~. Of course, when the paper with the finished spectrogram is removed from the drum and flattened out, the spiral appears cut into two hundred lines. These will be almost horizontal, with but a very slight inclination, the end of each on the right margin lying only $\frac{1}{200}$ of two inches (the total width of the graph divided by the total number of runs), that is, 0.01 inch higher than its beginning on the left

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

margin. This small deviation from the horizontal is not noticeable. Since all two-hundred lines are inclined at the same angle they are parallel.

The stylus, unlike that of a register-thermometer or barometer, does not actually register its whole course upon the paper but leaves a trace only at the moment when and at the frequency level where it is activated. The spectrogram can therefore scarcely consist of two-hundred slightly off-horizontal parallel lines. Even if one can produce a tone which is thus spectrographically visualized (it would have to be a sustained one of 18~ fundamental pitch which contains all of the possible harmonics up to 3600~), it is surely not a sample of speech lasting 2.3 seconds. In normal speech no single sound is so long sustained immutably, nor do we know of one of 18~ fundamental and two-hundred harmonics to 3600~. Normally, then, the stylus is excited intermittently when sufficient power is present at the frequencies which are scanned at this particular moment. Such power may be present for a given span of time, and it would be accordingly recorded as a line on the horizontal dimension, or as a mere dot if very brief, at a distance from the baseline equivalent to its frequency. (Compare Figs. 16, 17, 18.) In consideration of what I have said about the agility of the speech organs and their rapid positional changes for the production of many sounds following one another in quick succession in the normal stream of speech, even slow and deliberate speech, it can be expected that there will hardly be any sustained frequencies which might show as unbroken stylus traces of any length on a spectrogram of language. As I shall have occasion to point out, this is true even of vowels which last some measurable fraction of a second, because also within their duration, perceived by the ear as linguistically uniform, there occur acoustic variations. On a spectrogram of speech, therefore, continuous stylus markings will be virtually absent.

When discussing Figs. 17 and 18 I pointed out that the curves were drawn by hand, and I promised that this would not be necessary in actual machine-made spectrograms. In fact,

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

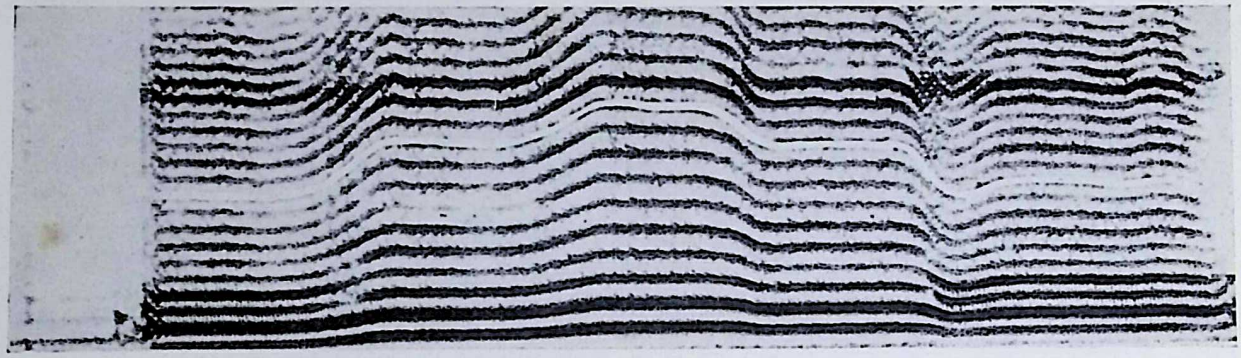


Fig. 21. Spectrogram, narrow band, *do-mi-sol-mi-do*, sung on vowel [i].

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

if Fig. 19 did not represent a tone sustained at the same pitch but a legato gliding melody, the horizontal lines would appear deflected in curves but they would still give the appearance of unbroken continuity. This effect is shown on *Fig. 21,* which is a spectrogram of the tune *do-mi-sol-mi-do.* The singer enunciated the vowel [i] throughout. The curves represent the harmonics of the legato melody, their rise and fall displaying the actual rise and fall of the pitch, in the manner of the hand-drawn lines on Figs. 17 and 18. For better visibility, the tenth harmonic is traced in white.

We shall soon see that this rendering of harmonics on the spectrogram is the particular effect of a narrow filter, for example of 45~ or 50~, which disappears when we increase the band width of the filter. The middle graph of Fig. 19 was made with a 300~ filter; and although it represents exactly the same tone as does the upper graph, produced with a 50~ filter, the visual result is wholly different. Such a change, or combination of spectrograms, may be desirable, and I shall presently explain its purpose.

I have already noted that the stylus cannot actually draw as continued lines such harmonics as are visible on Fig. 21, because the prescribed and inevitable path of the stylus, whether registering or not, is the spiral, which resolves, once the paper is removed from the cylinder, into two-hundred parallel lines. The only continuous line which a stylus can mark, when activated by rising or falling pitch, is a fraction of the spiral, or a fraction of one of the two-hundred lines at each revolution of the drum. But it cannot in one of its runs of 2.3 seconds leave its prescribed groove and move up or down traversing any number of frequencies. The curves of Fig. 21 must be, it follows, an optical illusion, albeit a most felicitous and instructive one, and we may ask ourselves just what produces this effect.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XV

18~ FILTER

If our spectrograph were provided with an ideal filter which allows but one frequency to pass, the stylus would react only to one single frequency, or, if there were a variable analyzing filter, to one frequency at each playback of the message, two-hundred in all in our setup. This means that on each playing of the tape only one frequency would be filtered out and that the stylus connected with the filter could indicate the presence of but one single frequency while all neighboring frequencies would pass unnoticed and unregistered. Since the filtering runs are 18~ apart, swaths 17~ wide would remain ignored and frequencies present therein would remain undetected. As far as the spectrograph and the spectrogram are concerned, such frequencies would simply not exist. But a filter of a certain band width gathers in all frequencies of sufficient power within its width (and the sensitivity to power can be adjusted by the experimenter). Hence an arrangement can be made by which *any* frequency, wherever it occurs, and at whatever moment, can be gathered in and is permitted to transmit its impulse to the stylus for registration. Furthermore, damped tones with subsidiary frequencies of some significance above and below the peak frequency of the harmonic, such as do occur in speech sounds, are thereby more accurately reported, inasmuch as a filter of a certain width can gather in several or all of the subsidiary frequencies at a given moment. This will have a concomitant effect on the total voltage by which the stylus is activated and a commensurate effect on the registration it performs. In other words a damped filter may enable us, (a) to cover the whole gamut of frequencies to be analyzed (in spectrography generally 0–3600~), (b) to obtain a truer visual representation of the acoustic effect of damping.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Let us examine all this in some detail, with special consideration of the purpose and the results of the use of filters with various widths.

Suppose that the spectrograph is scanning, at a given moment of its two-hundred runs, the frequency 900~, that is to say, it is filtering out this particular frequency from the total noise of the input. Let us further suppose that at this moment there is no 900~ component present in the tone, but a very important 909~ harmonic (the ninth of 101~, the voice pitch of an adult male speaker). No matter how powerful, that is, no matter how acoustically and linguistically significant this harmonic, an undamped filter that is set to 900~ cannot reach and cannot pass a 909~ frequency, hence there will be no registration of it by the stylus. The next scanning run of the machine will be 18~ higher; at the moment of the sample where in the previous run the filter and stylus were set to operate at 900~, they will now be 18~ farther on, at 918~. Obviously, at this run the harmonic of 909~ will be missed again by the undamped filter, which now is capable of passing 918~ but nothing else. Consequently those harmonics of a 101~ fundamental which, at the proper moment, do not occur at multiples of 18~ will not be registered, regardless of their power. The smallest common multiple of 18 and 101 being 1818, no harmonic of 101~ (occurring at the moment in question) below 1818~ can be registered on a spectrograph utilizing an undamped filter, nor can, of course, the 101~ fundamental itself. No valid visualization of the sample can result from such a scanning and filtering arrangement.

If the 909~ frequency is the peak of a damped wave, then it is possible that at 900~ there is strong enough a subsidiary frequency to be registered by the stylus, provided always that it occurs at the proper moment. In such a case, while it is not the 909~ harmonic itself that is registered, its presence would at least be acknowledged on the spectrogram through one of its subsidiary frequencies, rather than wholly ignored. But it should be noted that the event of the registration depends here

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

on the amount of damping: if the spectrum curve is very steep, a frequency of 900~, though only 9~ removed from the peak, may be so low in power that it is not strong enough to get itself registered, although in terms of frequency it passes a 900~ ideal filter. One may of course remedy this by increasing the power sensitivity of the filter, but in doing so one risks, for the sake of catching some subsidiary frequencies because they betray the presence of important peaks of power, to allow also a vast number of completely unimportant low-powered frequencies to clutter up one's graph, including possibly the ever-present and totally irrelevant frictional noises of the vocal apparatus. Thus many frequencies would intrude on the spectrogram to a degree far beyond their importance. It is therefore not advisable to attempt catching subsidiary frequencies for the purpose of having them take the place, at least graphically, of harmonic peaks. Consequently, even a damped wave with a spread of frequencies around a peak is not optimally analyzed by an ideal filter. Quite the contrary.

A compromise solution would be to retain the variable ideal filter but to have the machine make scanning trips at intervals of not 18~ but 1~. In this manner one would be assured of gathering all frequencies that occur, and one could arrange the power sensitivity of the machine in such a way that it does not register power below a certain minimum. But this would necessitate 3600 play-backs of the recorded sample of speech, and 3600 scanning trips (or more if the spectrogram is to go beyond 3600~); it would require a very wide strip of paper to accommodate all the registrations with good visibility, hence an increase in the size of the spectrograph itself; and the analysis of 2.3 seconds of sound would take more than two hours. (The machine can be accelerated but not beyond a certain limit for physical reasons.)

Rather than construct either an inaccurate or an unwieldy, slow machine, we have recourse to the damped filter of a certain band width. Let us assume, for the sake of illustration, that the spectrograph is now equipped with a filter 18~ wide.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

This width is normally not practical, but I choose it so that the sweep of the filter covers exactly the interval between the scanning runs. If now the machine is scanning at 900~, the filter is set to gather also all frequencies, if sufficiently powerful, from 9~ below to 9~ above 900~, that is, between 891~ and 909~. And now, of course, power present at 909~, the ninth harmonic of a 101~ fundamental, does leave its mark upon the paper regardless of whether the tone is damped or undamped. Since at the next run the machine is set, at the same moment of the playback, at 918~, and since now it sweeps again a width of 9~ above and 9~ below 918~, all frequencies between 909~ and 927~ can be passed, then those, at the 936~ position, between 927~ and 945~, and so on. Therefore, with a machine that is set for two-hundred scanning runs at intervals of 18~ and that is equipped with a variable filter 18~ wide and moving upwards at the same speed as the stylus, each frequency between zero and 3600~ is registered once. (The reason that the 909~, or any other in a similar time-frequency position, is not registered twice, as the highest frequency of the 900~ run and the lowest of the 918~ run, is of course that the stylus trace is a continued spiral, hence the end of the 900~ and the beginning of the 918~ run are one and the same geometric point.) No matter what the frequency of the fundamental at a given point, every one of its harmonics will be filtered and registered, and so will all sufficiently and significantly powerful subsidiary frequencies present on either side of the peak in a damped tone. (The threshold of sufficience and significance of power is determined by the experimenter who can adjust the machine to suit his purpose.)

In a machine running as I described it, a frequency of 909~ will be caught somewhere on the run of the stylus from 900~ to 918~. If it occurs exactly at the moment (on the horizontal dimension of the graph) when the stylus is passing through the 909~ spot (on the vertical dimension of the graph) while on its way from 900~ to 918~, then the registration will represent in terms of both time and frequency the exact physical truth.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

But if the 909~ frequency occurs at a moment when the stylus is either above or below the 909~ level, it will be registered not at the 909~ but at some slightly lower or higher level, somewhere between 900~ and 918~. Hence the registration, though it does take place at the proper point in time, may, but need not, take place at the proper frequency level. This inconvenience, however, is obviated for all practical purposes by the true dimensions of the spectrogram: a displacement of 9~, which would be maximal for the example just discussed, amounts to only 0.005 inch along the vertical, which is an imperceptible distance.

Fig. 22 (which is not in scale) illustrates the preceding discussion. The lines from left to right represent stylus runs along the graph paper. The crosses show the true locations, in time and frequency, of power present (at 909~) in the sample that is being analyzed. (It is immaterial just now whether we are dealing with a sustained tone at 909~ or with some kind of melody that touches 909~ three times.) The small circles indicate the location of the stylus, on one run, at the moments when power at 909~ is present. In part A, the stylus, coupled to an undamped, ideal filter, can register only the power present, if any, at the moment and at the frequency level at which is located, that is, about 903~. The 909~ component occurring at this moment passes therefore unnoticed and unregistered, no matter how powerful it is, because it lies at the wrong frequency level. That is to say, in the present setup the stylus can make only such registrations as render the physical truth in terms of both time and frequency. This fortunate coincidence takes place in part B of Fig. 22, where power present at 909~ and stylus position at 909~ concide. In part C the stylus is now coupled, not with an ideal, but with a damped filter 18~ wide, whose sweep is indicated by cross-hatching and which registers everything (of sufficient power) 9~ above and 9~ below the frequency level which it, together with the synchronized stylus, traverses at a given point in time. But the stylus performs the registration at the frequency at

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

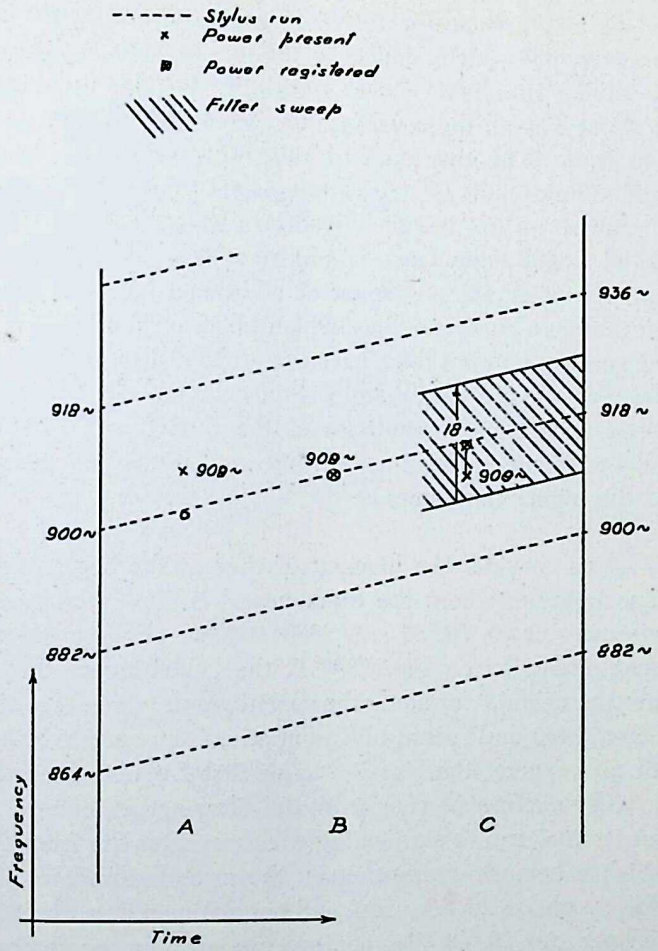


Fig. 22. Effect of 18~ filter width.

which it hovers at the moment and from which it cannot deviate along the vertical, and not at the true frequency level where the power is actually present. Hence the 909~ power is now registered on the present (the fiftieth) scanning run of the stylus—but at the level of about 914~.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Let us now assume that we are dealing with a legato kind of fundamental which, similar to the one of a tone produced by a sliding trombone, moves from 20~ to 25~ and finally ends up at 30~ in the course of 2.3 seconds, the length of the spectrogram. The overtones of this fundamental lie, at the corresponding points of time, at 40~–50~–60~, 60~–75~–90~, and so on up, in 120 fully drawn curves or 120 straight lines (on a 3600 wide spectrogram), to 2400~–3000~–3600~. (There are of course a number of additional not fully drawn but incomplete curves or lines which begin on the left margin of the spectrogram but then leave the graph at its upper margin before reaching the right margin. I have already explained why the overtones, though multiples of the fundamental, are not parallel to it and to one another, but increase in steepness toward the higher harmonics.)

Now let us consider the moment of time at the beginning of the spectrogram, where the fundamental is 20~, which gives us overtones at 40, 60, 80 .... 2380, 2400~. (Follow the subsequent discussion on *Fig. 23.*) At the beginning of the run around the cylinder, or along the spectrograph paper, the stylus will be at zero and record all frequencies from zero to half its width above zero, that is, 9~. The first frequency present, 20~, will therefore be caught by the filter not at its zero run but at its 18~ run (I am naming each run after the frequency at which it begins), and not again; the second harmonic, 40~, will be caught at the 36~ run, and not again, and so on. Since the frequencies 20, 40 60~ etc. are present only momentarily, our tone being a sliding one that does not dwell at any pitch, we shall obtain along the left edge of the spectrogram a vertical series of dots (180 of them, to be exact), distant 20~ from each other.

Next we move to the point where the fundamental is at 25~. Here, too, exactly for the same reasons, we obtain a vertical series of dots, 25~ distant from one another (though now there is room for but 150 of them). At the fundamental frequency

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

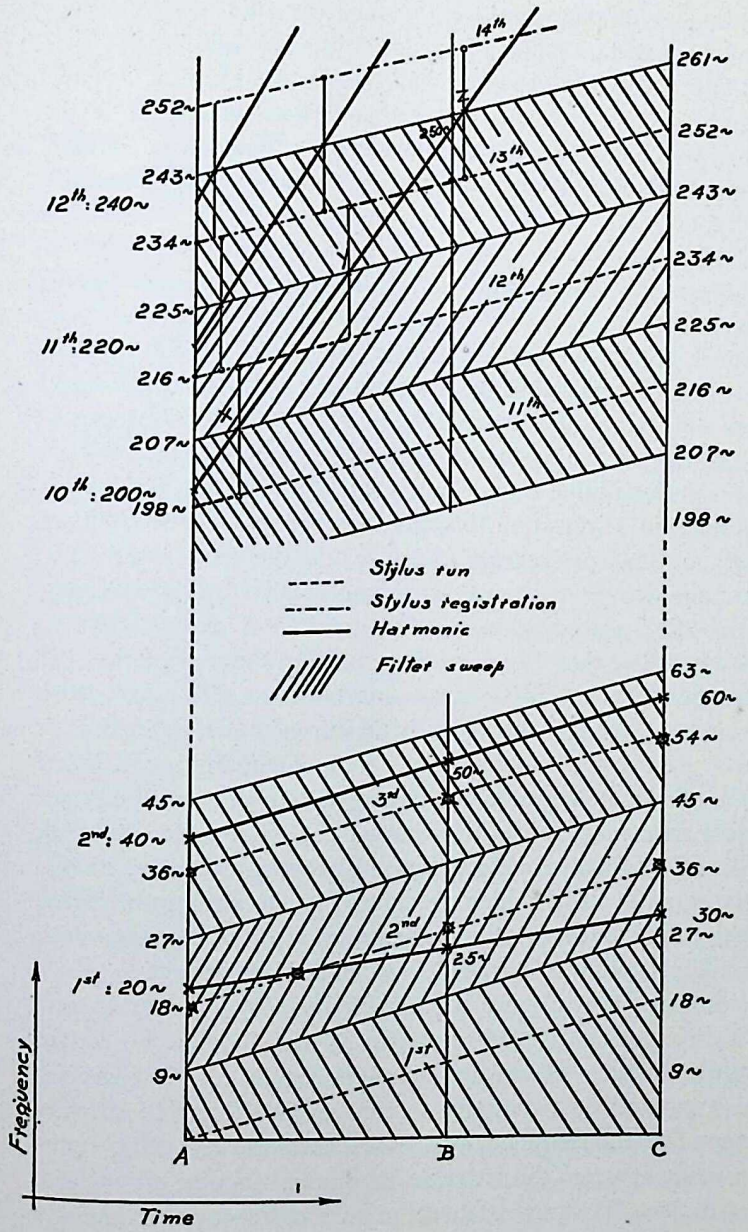


Fig. 23. 18~ filter, legato rising tone.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

of 30~ we obtain, therefore, dots at intervals of 30~ (120 of them). Up to this point in my description we have once more the fever curve type of chart in which we may draw by hand connecting lines between points that belong to the same harmonic, hoping, with more or less certainty depending on other information, that the resultant lines represent the acoustic reality.

On the spectrogram, however, this is not necessary if we consider what the machine accomplishes between the three points I arbitrarily selected for illustration and measurement. We know, after all, that the stylus moves steadily, and we also know that the filter is continually receptive to frequencies present within its reach. Now the 20~ frequency is, in the case of an upgliding sound, present only a moment. The stylus continues on its run from 18~ to 36~, moving across the whole length of the spectrogram paper, while the 18~ wide filter continues sweeping concurrently a band which at the beginning of the 18~ run covers a swath between 9~ and 27~, then 10~ and 28~, then 11~ and 29~, until at the end of the run, with the stylus at 36~, it sweeps between 27~ and 45~. Throughout this run the stylus is, in our example, activated by power present at 20~ only once, on the left margin, and never again. It is, however, activated again on the same run, at certain moments, by the presence of the frequencies between 20~ and 30~ (this is the progress of the fundamental on the spectrogram), all of which are bound to lie within the reach of the filter. But since the stylus moves upward together with the filter from 18~ to 36~, and since the frequency curve moves concurrently upward from 20~ to 30~, that is to say, at a different rate, the frequencies at which the stylus leaves a mark do not necessarily represent exactly the frequencies which activate it, as we have seen earlier (Fig. 22). It may happen that the frequency of 25~ excitates the recording stylus at a moment when the latter, in its gradual regular climb from 18~ to 36~, has only reached the level of 23~, or has already

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

progressed to 27~. Indeed, if 20~ was present at the very beginning of the run it got registered at the level of 18~; and the 30~ prevailing at the end of the run is marked on the 36~ level on the right hand margin of the spectrogram. I have already commented on our freedom to ignore this distortion because of the infinitesimal distances involved on the spectrogram.

Now immediately after registering the 20~ on the left margin the stylus moves slightly upward. If after a fraction of a second there is still power present within reach of the filter the stylus remains activated, and it continues to burn its mark upon the paper as long as power remains within reach. On our example (20~ to 30~ gradually rising tune, and an 18~ filter on the 18~ run) there is power present throughout the run of the stylus from 18~ to 36~, with the 18~ wide filter sweeping a width of frequencies from between 9~ and 27~ on the left to 27~ and 45~ on the right (cf. Fig. 23, lower half, on which the points *A*, *B*, and *C* indicate the moments at which the fundamental is 20~, 25~, and 30~, respectively). Therefore, the stylus is activated throughout its second run, from 18~ to 36~, and burns a continuous trace on the paper. This trace is a straight line, one two-hundredth of the entire spiral which the stylus describes.

A similar registration, though 18~ higher, occurs on the third stylus run, from 36~ to 54~, with the entire course of the harmonic still lying throughout within the breadth of the third filter sweep.

Let us now examine the tenth harmonic of the sliding tone, where the corresponding points *A*, *B*, and *C* of the time scale come to lie at 200~, 250~, and 300~. The 200~ at the left margin will be gathered by our 18~ filter on the eleventh run. This run begins at 198~ on the left, where the filter sweeps between 189~ and 207~, and ends at 216~ on the right, where the filter sweeps between 207~ and 225~. In contrast to the second run, where the entire first harmonic was gathered in by the filter on one trip, on this eleventh run the filter no longer

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

reaches the 250~ and 300~ points of the tenth harmonic, which means that the stylus now is *not* activated throughout this run and that the tenth harmonic cannot be registered in one long uninterrupted straight line.

How long, then, after having registered the 200~ frequency on the left margin, does the stylus continue marking? It does so until the frequencies present leave the reach of the filter. Somewhere the harmonic will leave the 18~ wide strip (indicated by cross-hatching on my sketch) which is swept by the filter. This occurs at the point *X* on Fig. 23, upper half. However, since in the 18~–18~ arrangement which we employ (runs at intervals of 18~, and filter 18~ wide) the upper edge of one sweep coincides with the lower edge of the next sweep, the stylus will resume marking on the next run exactly on the point in time where it left off in the previous run, though at a point, on my sketch, 18~ vertically above the end of the previous marking. On the twelfth run the stylus is activated again as long as the frequencies are within the filter pass band, that is, to point *Y*, then on the thirteenth run to *Z*, and so forth. The total registration which the stylus draws is, then, in reality no line at all but a series of short, slightly inclined lines, each a little higher (or lower, if the pitch of the sliding tune falls) than the preceding. But just as the deviation from the horizontal is so small as to be visually imperceptible, so the step-like effect on my strongly distorted drawing is reduced to imperceptible proportions on the real spectrogram: the distance of 18~ amounts, on the 2 inch wide spectrogram, to a mere 0.01 inches. The impression of a continuous curve, the optical illusion of which I spoke in discussing Fig. 21, is thereby explained.

A number of deductions can be made from Fig. 23. (1) If the tune, or the sound, has so flat a fundamental (in our example 20~–25~–30~) as can be registered by an 18~ wide filter in one single stylus run, with the result that the stylus remains activated throughout its run and marks a trace running straight through from 18~ on the left to 36~ on the right, then this

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

registration does not tell us anything, at the level of the fundamental, about the true frequency curve. A straight 20~ evenly sustained pitch throughout the 2.3 seconds, or a 15~–25~ curve, or a 21~–30~–23~ curve, or any other harmonic which is capable of activating the stylus without interruption, would have produced the same spectrographic result. (2) But since the steepness of the line increases of necessity in the upper harmonics, any such ambiguous cases may be resolved by reading off the information concerning the true pitch of the fundamental at the level of sufficiently high harmonics. (3) If the tune has steep enough a fundamental, say, 20~–60~–90~, then the curve effect which more nearly approximates the true pitch becomes visible even at the fundamental, because an 18~ filter cannot reach beyond 45~ on the 18~ to 36~ run, hence will fail to catch the frequencies of this steeply rising fundamental soon after the beginning of the run. (4) The steeper the rise or fall of the harmonic, the narrower the component steps on the level of each stylus run. They may indeed shrink to the size of points, so that the resultant 'line' on the spectrogram is actually composed of many dots, of which each lies slightly higher (or lower) than the preceding one, and to the right of it. The optical illusion of a continued curve may thereby be further enhanced. (This will indeed be the case more often than not in spectrograms of speech, where frequencies vary rapidly and steeply.) Since the steepness is greater in the higher harmonics, it follows that, in the analysis of the same sound, the higher the harmonic the more will its spectrographic contour be composed of dots rather than steps. (5) Consequently, if we deal with a higher fundamental than the one on Fig. 23, we reach the dot feature at lower harmonics. Such is the case in speech, where the fundamental of a male voice lies not around 25~ or 30~, but nearer 100~ and above.

This is the way, then, in which the spectrograph builds up piece by piece the seeming curves which indicate the pitch contours of both fundamental and harmonics, thus eliminating the necessity of connecting points on the same curve by hand.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

We owe this to the ingenious arrangement of a spiraling stylus coupled with a filter of a width greater than zero (minimally 18~ for reasons already stated).

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XVI

50~, 300~ FILTER

What happens if we install a filter with a pass band wider than 18~? This is illustrated in *Fig. 24* for a 50~ filter. I have again chosen a fundamental gradually rising from 20~ through 25~ to 30~, but traced it around the seventh (108~–126~), eighth (126~–144~), ninth (144~–162~), tenth 162~–180~), and eleventh (180~–208~) stylus runs. The harmonics concerned are the 5th, 6th, 7th, 8th, and 9th. The outstanding feature of this diagram is the overlapping of the areas swept by successive filtering runs. This is of course due to the fact that the pass band of 50~ is wider than the 18~ distance of the stylus runs: a frequency of 140~, for example, on the left hand margin of the spectrogram, activates the stylus on its 126~ run (when the filter reaches as high as 126~ + 25~ = 151~), on its 144~ run (144~ – 25~ = 119~), and on its 162~ run (162~ – 25~ = 137~), or three times altogether. Other frequencies at other points may be registered only twice. But on the present 18~–50~ setup no frequency present is registered less than twice. This overlapping is shown by areas of double and triple cross-hatching on my illustration, each hatching indicating one filter sweep.

Since the filter width (50~) is, on the present example, also greater than the distance between the harmonics, it may occur that the filter passes, on one and the same run, more than one harmonic. For example, at the beginning of the 144~ run, where the filter reaches as low as 119~ and as high as 169~, the 120~, the 140~, and the 160~ overtone of the 20~ fundamental are caught. A stylus run may therefore register, over its whole length or any part of it, one or more entire or one or more partial harmonics, or any combination of total or partial

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

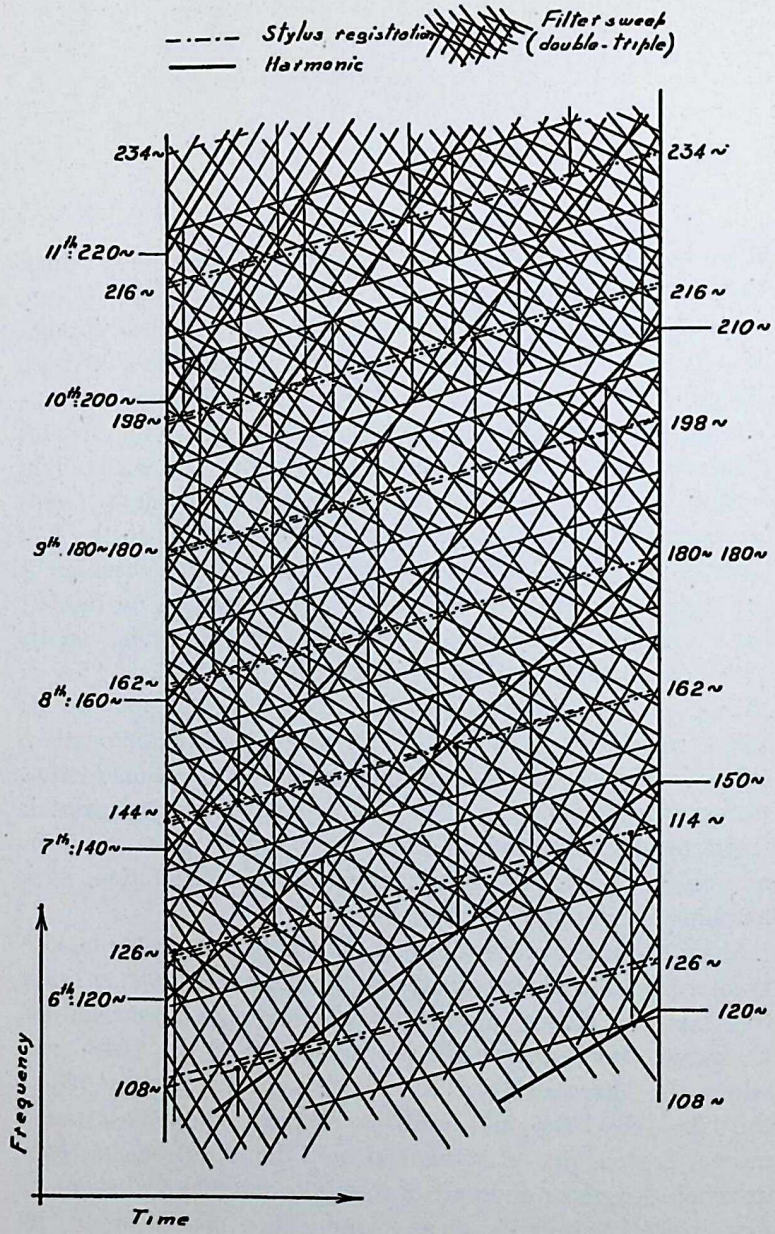


Fig. 24. 50~ filter, legato rising tone.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

harmonics. For example, the 144~ run (9th filter sweep) registers, on my graph, the entire 6th harmonic, more than the first half of the 7th, about the first sixth part of the 8th, and the last third of the 5th. (This can easily be verified by blocking off by means of two rulers or sheets of paper, everything above and below the 9th filter sweep: whatever harmonics remain in the visible part are registered on the 144~ run.) The 180~ run, however, (11th filter sweep) gathers in no harmonic in its entirety, only parts of the 6th, 7th, 8th and 9th. I have indicated, on Fig. 24, this gathering-in of more than one harmonic by multiple tracings along the appropriate stretches of the stylus registration. If the stylus is tuned to react to variations in power input, it leaves along such stretches darker tracings due to greater accumulated voltage caused by those multiple impulses. But note that in spectrograms of speech, where the harmonics lie 100~ and more (equivalent to the frequency of the fundamental) apart, a 50~ wide filter will not be able to catch two harmonics simultaneously.

With a 50~ filter, the frequency of 909~, which we considered before, no matter when it occurs on the 2.3 seconds long sample, can escape registration no more than it did on the 18~ filter spectrogram. (See *Fig. 25.*) Indeed, it will be registered more than once. If it occurs at the left margin of the spectrogram, at time zero, it will lie within the 50~ filter sweep on the following runs (named as usually after their initial frequency): 900~ and 918~; if it occurs on the right hand margin of the spectrogram, at time 2.3 seconds, it will be registered on the following runs (still named after the initial frequency; but their final frequency at the end of the run is 18 higher): 882~ and 900~; if it occurs somewhere between zero and 2.3 seconds, it will also be caught twice on two of the three runs, 882~, 900~, and 918~, and three times, once on each run, between points *A* and *B*, that is, where it lies in an area of triple hatching. To put it differently, I may also say that, in the analysis of a 909~ sustained tone by means of a 50~ band width, the 900~ stylus run is activated throughout

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

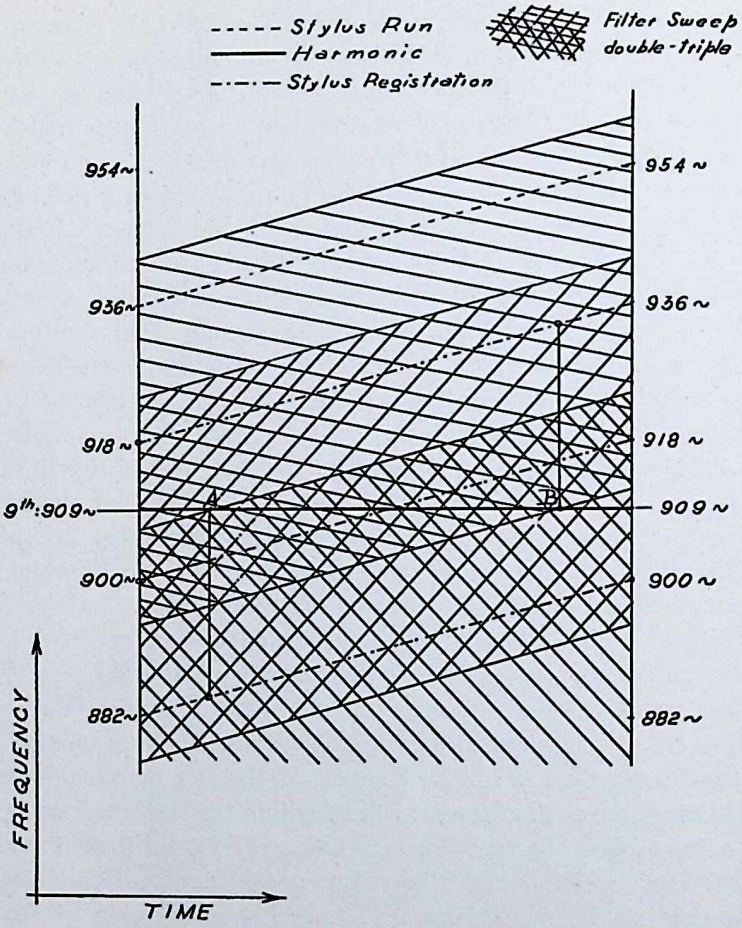


Fig. 25. 50~ filter, 909~ harmonic.

its length, the 882~ run in its last eight ninths (approximately), and the 918~ run in its first eight ninths (approximately).

The difference between an 18~ filter and a 50~ filter spectrogram lies, then, in the overlapping of filter sweeps, producing more than one registration on successive levels for certain pieces of each of the harmonics, and in the possible gathering-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

in of more than one harmonic on each stylus run. Now the more frequencies the filter can gather, the greater the power that impinges upon the stylus, and the greater the voltage with which the stylus is charged, which shows itself in greater blackness of the trace it burns upon the paper. It follows that one and the same stylus run may show different degrees of darkness, depending upon the number of frequencies that are being filtered out and registered at a given moment.

There is another aspect in which the 50~ spectrogram differs from an 18~ one. Since successive registrations occur 0.01 inches apart from one another on the vertical, this being the distance between each two of the two-hundred stylus runs, two successive registrations of the same frequency are therefore 0.01 inches distant from one another on the vertical. Of course, this small interval will scarcely be visible as such, but two dots thus drawn by the stylus do appear to the naked eye as a single thicker dot or short line, measuring 0.02 or 0.03 inches on the vertical in the case of two or three registrations, respectively.

Note, however, that the simultaneous passing through the filter, on the same run, of two harmonics can contribute, not to greater size, but only to greater blackness of the mark because of the higher voltage impetus received by the stylus. In general, therefore, we should distinguish the degree of *blackness* of the registration which is due to the amount of power, that is, the number and collective power of *various* frequencies present at any moment of a *single* run, from the *breadth* of the registration which is due to the repeated passing of the *same* frequency at the same moment of *successive* runs. These two parameters together are responsible for the ultimate aspect of the registration.

Both breadth and blackness of the registration, be it line or dot, of the same sound or part of a sound can be increased at will by widening the analyzing filter (see below), for this brings about the conditions of frequency-gathering necessary for more intensive (blacker) and extensive (wider) markings.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Note also that, since in speech we are dealing almost invariably with damped sounds, that is, with a number of frequencies clustering closely about the peak frequencies constituting the harmonics, the simultaneous passing of several frequencies at any moment, hence a voltage greater than that caused by the harmonic alone, hence a fair amount of blackness, are assured, as long as a damped filter is employed.

If we installed a filter 3600~ wide, that is, as wide as the whole spectrogram, the result would be that the stylus would on each run be activated by any frequency present between zero and 3600~. (It would fail to be activated only when there is no frequency at all present in the total range. And this happens only when complete silence prevails.) On the spectrogram this condition would cause two-hundred parallel lines 0.01 inches apart, which would simply show as a total blackening of the paper, interrupted only by vertical swaths across its total height indicating silence, if such occurred. Obviously, such a filtering arrangement is worthless because it is not discriminating enough.

If we install a 300~ wide analyzing filter, we obtain, as a sketch like Fig. 24 would show, for the 909~ overtone of a 101~ fundamental not two or three but up to sixteen successive registrations, that is, a line up to 0.16 inch thick, extending 0.08 inch above and equally far below at some point or stretch of the 900~ run of the stylus. Also, a filter 300~ wide is able to gather two or three neighboring harmonics of a 101~ tone, and the 900~ run can catch not only the 909~ but also the 808~ and the 1010~ harmonics, if they are present, because the filter reaches 150~ below and 150~ above its position at a given moment of the 900~ run. The result of this is, as usually, greater darkness of the tracings.

Now if, in our example, a harmonic at 808~ is actually present and of sufficient power to cause activation of the stylus, then there occurs a registration of it not only on the 900~ run just discussed, but also another registration on the 792~ run, extending again 0.08 inch above and 0.08 inch below it. Since

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the 792~ and the 900~ runs are, on the spectrogram, 108~ or 0.06 inch apart, registrations 0.08 inch wide coming from above and below will overlap. In this manner the clear and visible separation of the 808~ and 909~ harmonics is blotted out. An experienced spectrographer is not misled by this since he knows that a 300~ filter cannot produce, from the activation of the stylus in just one run, either markings as thick as those due to overlapping registrations of several runs, or, in case of the momentary occurrence of a frequency, a 'dot' that is now really more like a vertical line instead of a real dot. But he might have to do some calculating to find out just what exactly this wide band, or this vertical line, does represent; and the technically untrained viewer of the spectrogram derives from it, at first sight, little significant information concerning the number and distribution of harmonics present. If, therefore, one wishes to resolve spectrographically the harmonic composition of a tone, one must not choose a filter whose width is greater than the distance between the harmonics.

A tone like the glottal tone has many powerful, acoustically and linguistically important harmonics. Since the fundamental frequency of human speech is rarely higher than 300~, it follows that a 300~ wide filter will produce a spectrogram of speech in which overlapping and lack of clarity are almost inevitable. If one considers also the effect of damping in the supra-glottal cavities, with its many subsidiary frequencies, and the consequent darkening of the spectrographic markings due to the high voltage caused by many frequencies activating simultaneously the stylus at any given moment, one might expect that a spectrogram of a speech sound produced with a 300~ filter would show merely vast areas of deep blackness from which details can scarcely be gleaned.

The following chapter explains how this threatening deficiency of the 300~ broad band spectrogram is not only overcome but also converted into an asset.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XVII

FORMANTS

As we have seen (Chapter VII above), the *practical* width of a damped resonator-filter, like that of a damped sonator, as distinguished from its *theoretical* width which is infinite, is defined by its half-power points. This means, then, that by filtering a sound through a filter of a given width we not only eliminate from the spectrogram frequencies outside the filter's resonance curve in terms of frequency, but we also ignore frequencies below a certain amount of power.

I am illustrating by means of several resonance curves (compare Fig. 13) the relation of filter and sound in *Fig. 26* A and B, where one and the same filter is exposed to a variety of tones. The broken line marked *f*, represents the resonance curve of the filter, and the solid lines marked *1, 2, 3, 4,* in ascending order of power, four different sounds. M, N, M_x, and N_x mark the frequency-power location of the half-way points of the filter: line x is the powerpass limit, the distance M—N defines the filter width of n∼. Now if a filter of this type is installed in a spectrograph, any frequency that is to pass it and that is to be registered by the stylus of the spectrograph must fulfill two interdependent conditions arising from the nature of the filter: first, it must be a passable frequency, that is, it must lie within the resonance band of the filter; second, it must be of a power above that of the half-power points of the same filter. Consequently, only frequencies between M and N (frequency) and above M_x and N_x (power) will be spectrographically registered. In Fig. 26 A, therefore, certain parts of curves *2, 3,* and *4,* which I have provided with short cross strokes, qualify in terms of both frequency and power; curve *1,* although containing suitable frequencies, is throughout of insufficient power.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

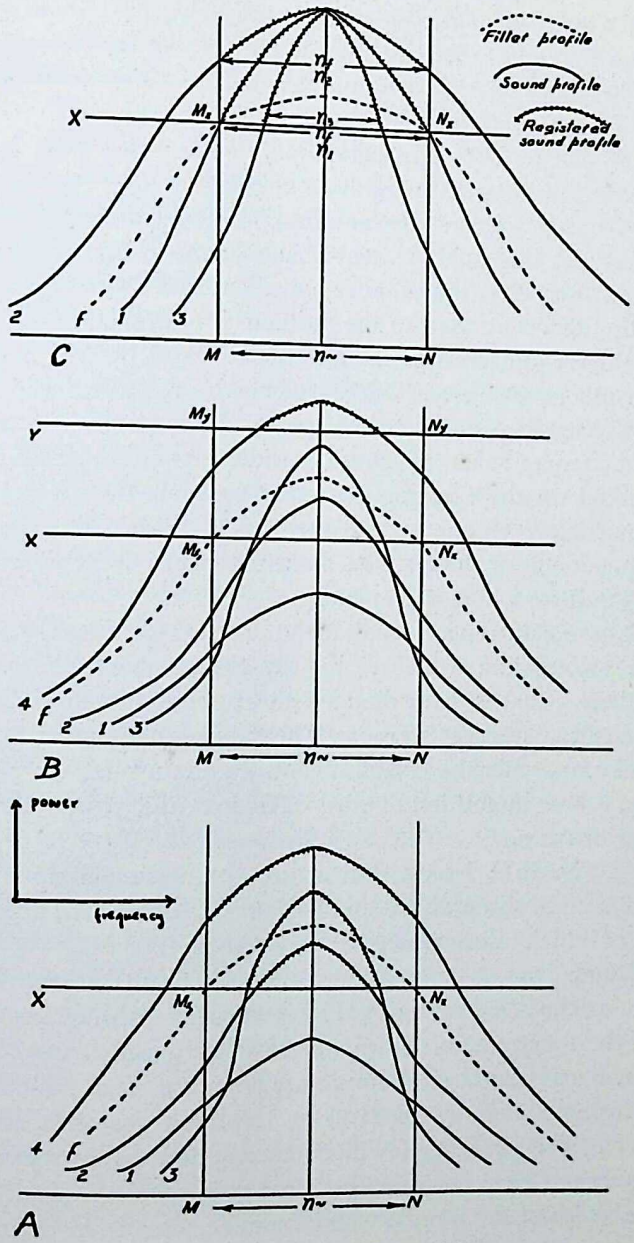


Fig. 26. Damped tones and damped filters.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

A curve of sufficient power but unsuitable frequencies would lie to the right or left of the curves on our sketch, its peak so far away from the peak of the sketch that its own M–N distance would not overlap with that of the sketch.

Of course, the spectrographer is not forced to accept the filter's conventional half-power points as the power pass limit for his spectrogram. He may find it useful either to raise or to lower the threshold without changing the band width of the filter. Fig. 26 B shows what happens to the same four curves if the line x is raised to the position y: now only the strongest of the four curves will activate the stylus of the spectrograph, the others, being too weak, will not be registered. Thus the experimenter can select those peaks of power on the spectrum of a complex sound which he considers important enough to be marked on the spectrogram, and weed out the others which may merely clutter up the graph. He can do this quite independently of the frequency characteristics and the spectrum of the filter which is employed.

I have said repeatedly that the darkness of the markings on the spectrogram will increase with the number of frequencies gathered, that is, with the amount of power present at a given moment of the registration. The process of filtering just described permits the establishment of certain relations between these three quantities: number of frequencies, power, and darkness of trace.

In Fig. 26 C, I have drawn three profiles of damped tones, *1*, *2*, and *3*, all of which are of the same power at the peak and all of which are registered in part when a filter with the profile f is employed. The question is now whether these three tones, having the same power peak, will register with equal darkness on the spectrogram. Since we are dealing with damped tones that have a number of subsidiary frequencies, of which a certain number are gathered in by the filter, power in excess of that of the peak itself will be transmitted to the stylus. The number of frequencies so gathered corresponds in Fig. 26 C to the width of the bundle whose baseline is n_f, n_1, n_2, n_3, respect-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

ively. The maximum width is of course that of the filter itself, n_f. The same maximum pertains also to tone *1* (n_1) and tone *2* (n_2), and also to all other tones whose spectrum profile cuts the line x at or outside of the half-power points M_x and N_x. Profiles cutting inside these points represent tones which contribute fewer frequencies. It follows from this that the darkness of spectrographic representation is dependent, not only on the height of the profile (power, loudness), but also on its shape (frequency composition, quality). That is to say, darker spectrograms may represent not only louder tones but also, for example, blunter tones (which have a flatter profile). This further reduces the quantitative exactness and trustworthiness of determination of power, or loudness, from the spectrogram. But certain advantages outweigh this disadvantage.

It will be recalled that the very complex glottal tone undergoes, upon entering the supra-glottal cavities, modifications of quality and power, though not of fundamental pitch, through resonance and reinforcement, owing to the size and configuration of the cavities. Each cavity causes that particular frequency of the glottal tone at which the cavity itself would sound, to resound. This produces a rearrangement of power among the frequencies present in the glottal tone, that is to say, a modulation of the quality of the original glottal tone. The spectrum of the modified glottal tone will therefore be quite different from that of the unmodified glottal tone that issues from the larynx.

It is known through experiments that the glottal tone, when vowels are articulated, is modified in such a way that, between the fundamental and 3600~, its spectrum profile attains certain peaks of power, and that among these there are two, three, or four (depending on the vowel—but three is the usual number that shows, or is made to show, on the spectrogram: see below) which are particularly high. *Fig. 27* shows a spectrum of this kind. (No attempt was made to represent any particular vowel. Cf. Figs. 11, 13, 19 bottom). Spectrographers and acousticians do not know for certain to what feature of oral resonance these

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

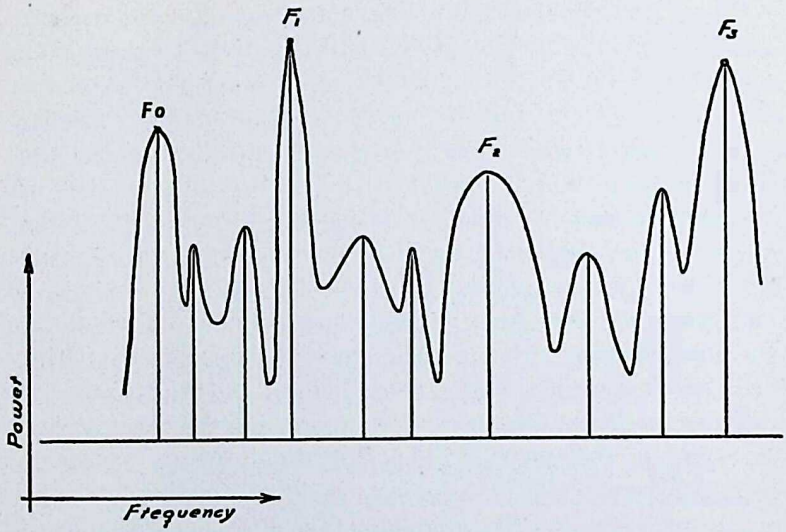


Fig. 27. Formants.

peaks are due. Some suspected, though erroneously, it now appears, that the first three above the glottal pitch peak (which is the lowest in terms of frequencies though not necessarily the weakest in terms of power) represented the resonances created by the three principal oral cavities constituted by the position of the tongue and the lips as shown in Fig. 15. The truth seems to be, rather, that all cavities are linked, or coupled, and that no single one is singly responsible for any one peak of power.

Filtering this kind of spectrum with several peaks is, in principle, not different from filtering one with but one peak, such as shown in Fig. 26. It is obvious, however, that a damped filter of a given width may now span and pass several rather than one peak at a time, at one scanning run, and that some of the peaks may now pass and register on several successive scanning runs. How many peaks pass the filter depends, naturally, again on the width of the filter and on its power pass. And for any given filter width the power pass may be varied so as to include the strongest, or the four strongest, or the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

twenty strongest peaks on the spectrum, depending on the kind of analysis one wishes to perform and on the number of peaks that one wants to have on the spectrogram. All this has been fully discussed in connection with Fig. 26.

In spectrography these peaks are referred to, for by now obvious reasons, as FORMANTS, numbered F_0, F_1, F_2, and so forth, from the lower frequencies toward the higher. (Some authors call the fundamental voice pitch, though it is not really a formant, F_1.)

Non-resonant sounds lack these peaks of power, or formants, since their articulation and phonetic quality are not primarily and significantly determined by resonance in the cavities.

The formants of (non-nasal) resonant speech sounds produced by an average male voice lie normally within the following frequency ranges:

F_1	150– 850~
F_2	500–2500~
F_3	1700–3200~
F_4	2500–4500~

The average female voice has formant frequencies about 17% higher than the male voice. This difference is due to the female's normally smaller total length of the vocal cavities from the glottis to the lips. Children's voices show even higher formant frequencies.

The fundamental pitch of the female voice is on the average about one octave higher than that of the male voice. Typical mean values are 220~ for females, and 120~ for males.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XVIII

RESONANCE BANDS

On broad band spectrograms the formants appear as so-called RESONANCE BANDS, or RESONANCE BARS. I shall now explain why and how they show in this particular form.

What filters of various widths can do, and how they do it, has already been described in detail. An important part of the procedure was to choose filters not only with regard to a certain pass band of frequency, measured in width of cycles, but also with regard to a pass band of power which determines the threshold of power sensitiveness of the machine. (Cf. the line *x* or *y* on Fig. 26). We may, therefore, employ for the spectrographing of sounds, in our case of the vowels, this kind of filter set at such a degree of sensitivity to power that the machine registers, for example, only the relatively high amplitudes which are characteristic of the formant spikes, whereas lower amplitudes are not registered. Consequently, if I choose an 18~ wide filter which on stylus runs 18~ apart may be capable of catching all frequencies present, I may at the same time prevent the indiscriminate registration of frequencies, of which many are linguistically irrelevant, by adjusting the power sensitivity of the spectrograph in such a way that it remains impervious to all but the three highest, or the five highest, or *n* highest spikes of amplitude. I could, for example, obtain a spectrogram which shows all those peaks of power which correspond to the harmonics. The result of such an analysis would be a spectrogram on which one discerns a number of wavy lines one above the other, each of which represents the frequency contour of one of the harmonics, and whose distance from one another on the vertical, the frequency dimension, corresponds, naturally, to the frequency of the fundamental itself.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Since not all harmonics are equally powerful and since, in a damped wave, the peaks are more or less steep according to the amplitude distribution of the subsidiary frequencies of each harmonic, the voltage transmitted to the stylus will vary, and so will, consequently, the darkness of the registration. On an 18∼—18∼ setup of the type discussed before, each frequency is gathered in and passed through the filter only once; hence all marks left by the stylus will be of the thickness of one registration, and the indication of the amount of power present will be visible only as greater or lesser blackness in dots or lines of equal width. Obviously such a visualization of amplitude, or power present, will not only be, as one would expect, non-quantifiable on the spectrogram, but also the degrees of blackness will be scarcely perceptible in dots or lines so thin as those drawn by the stylus.

As we have seen, there exists a device whereby the registration of each frequency that reaches the stylus can be multiplied: this we have accomplished by the use of filters wider than the distance between scanning runs. By using a 50∼ filter, for example, each frequency is registered twice or three times. If we now analyze a sound on a 50∼-18∼ setup and adjust the power pass in such a way that, as above, all harmonics will register, we obtain again the series of wavy lines on the spectrogram which represent the harmonics. But now each registration will be due to two or three impulses of the stylus 0.01 inch apart, which will appear, as I said before, as dots or lines of double or triple thickness. With this increase in dimension of the registration, visual differentiation of varied shadings of blackness also becomes easier. Now the identification of the principal powers among the harmonics, of the regions within the spectrum where power is transmitted most efficiently, that is, of those partials which contribute most to the quality of the sound at any given moment and which we called formants, becomes possible with greater certainty.

If we made a narrow band, 50∼ spectrogram of the English vowels as listed in Tab. I, we should obtain pictures resembling

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the upper graph of Fig. 19: the harmonics, at least all those strong enough to pass the filter, would show, if each vowel is briefly sustained as is the cello tone of Fig. 19, as horizontal lines. But this type of narrow band spectrogram, although it has its value as rendering the pitch contour for each of the component frequencies of sufficient power to have passed the filter and activated the stylus, is not apt to give us very good information on those parameters which are typical and typically different for each vowel phoneme, that is, the formants. But it is the formants we want to see, because the supra-glottal modifications of the complex glottal tone, which is composed of many harmonics, emphasize but few of them, thanks to resonance and reinforcement: it is only these few selected, resonated and reinforced, harmonics which give the speech sound its typical quality and which are therefore called formants. And it is only they one needs to isolate and make visible for the kind of visualization that attempts a distinctive picture for the distinctive sounds.

Of course, we could rig the power pass of the machine in such a way that only the frequencies of enough amplitude to qualify as formants are allowed to register. But, as one can easily see, this would deliver rather skimpy spectrograms, even with a 50~ filter, let alone with an 18~ filter. From these it would be difficult to learn the individuality of phonemes. The answer to the problem lies obviously in the use of a yet wider filter.

With a 300~ wide filter each frequency that shows on the spectrogram appears, thanks to the phenomenon of multiple registration, as a vertical line (as many times 0.01 inches high as there are registrations) if the excitation is momentary, or as a horizontal band of the same width if lasting; if several frequencies follow one another in a gliding legato manner, the result is a wavy band of blackness.

I mentioned toward the end of Chapter XIV that a 300~ filter spectrogram runs the danger of producing great fields of unintegrated greyness or blackness, and Chapter XV showed

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

how a proper installation of power-pass could prevent this. If, therefore, we do use a 300~ filter we must see to it that the power sensitivity of the machine is set at high enough a threshold so as to prevent all but the highest amplitudes from excitating the stylus. But since, indeed, in the spectrogram of a vowel all we want to see is what exactly makes this vowel different from the other vowels of the language, and since the difference in quality resides acoustically in the contribution of the strongest partial amplitudes and spectrographically in the formants, a spectrogram that shows mainly the formants—but shows them clearly and extensively, as does one produced with a wide rather than a narrow filter—renders a linguistically valid, instructive, and distinctive visualization.

A wide filter spectrogram which registers more than F_1, F_2 and F_3, the three principally interesting formants, may still be quite acceptable, as long as it is not really too cluttered with frequencies. For thanks to the multiple registration of many single frequencies and the simultaneous passing and registration of several frequencies, different shades of black or grey arranged in broad bands can now be differentiated from one another easily by the viewer: the blackest and widest bands on a spectrogram always represent the main formants, that is, those areas of concentrated frequencies where the filter passes and where the stylus records the greatest numbers of frequencies simultaneously, together with the greatest numbers of multiple successive registrations. But one may invariably omit the registration of glottal pitch (sometimes marked F_0): first, because glottal pitch is not linguistically significant; second, because, as I said before, the stylus-drawn pitch contour is unlikely to present faithfully the real pitch contour in the lower frequencies, so that, even if one wanted to know the latter, it would be better to read it off at the level of the higher harmonics.

The absolute position and the width of the formants themselves have little to do with the distinctive quality of the vowel; instead, they are related to the pattern of formant frequencies,

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

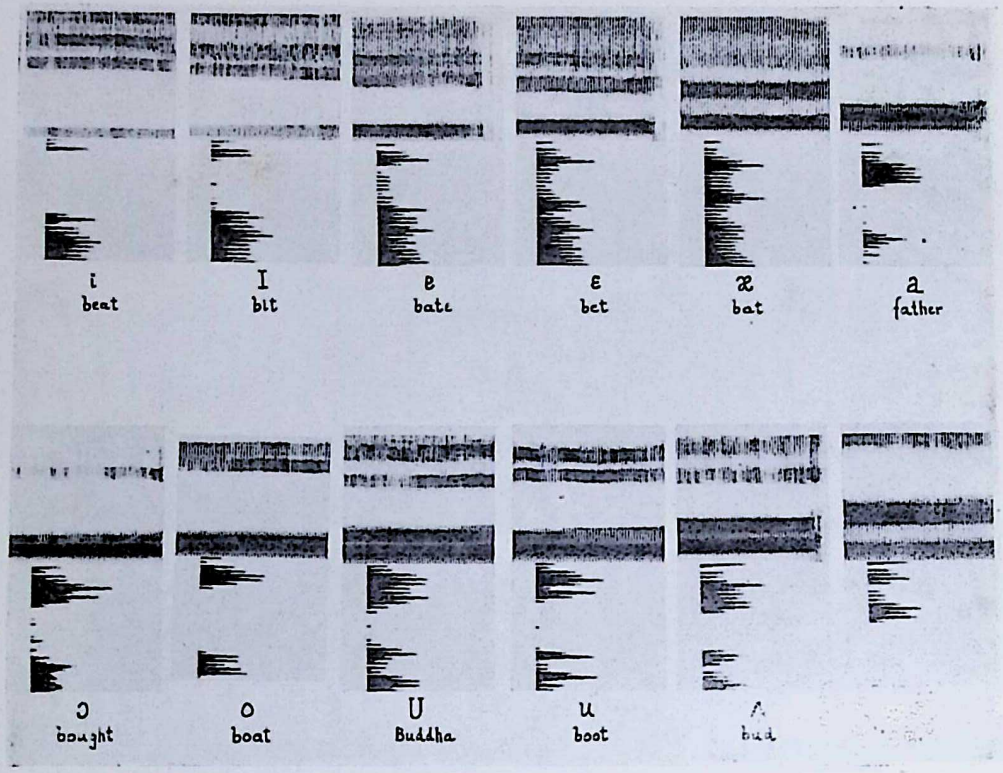


Fig. 28. Broad band spectrograms and amplitude sections of English vowels.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

that is to say, to articulation. In other words, absolute formant levels are determined by the fundamental pitch, formant width by the idiosyncrasies of individual articulation. But it is the relative position of the formants and their distribution which are correlated with the vowel quality.

The upper halves of the pictures in *Fig. 28* are 300~, or broad band, spectrograms of the vowels listed in Table I. Note that we have lost the resolution of the complex wave into its harmonics, but that we have gained a more easily visible and more useful showing of the significant formants through bands of blackness. These are called, as a feature of the spectrogram, (resonance) bands or (resonance) bars, the name being due to their representing the resonances created in the cavities by certain of the component frequencies of the glottal tone.

It should be mentioned that the spectrograph responsible for the spectrograms here shown has a 'high-shaping' setting, which provides the high frequencies, generally weak in speech, with a general emphasis. This means that amplitudes of the upper formants are emphasized more than naturally. But since we are looking, in such spectrograms of speech, only for the relative position of the formants, and not for their absolute power and position, high-shaping involves no significant distortion. (The spectrograph can also be set for low-shaping if necessary or desirable.)

The lower halves of the pictures of Fig. 28 are amplitude sections, or spectra, like those of Fig. 19. Since the vowels, or the portions of each vowel spectrographed, are sustained unchanged throughout their brief duration shown, hence exhibit horizontal and parallel resonance bands due to evenly sustained articulation, a spectrum snapshot taken at any point is a true and complete picture of the frequency-power composition of the vowel (cf. what I said concerning the first spectrum I showed, Fig. 11). From such amplitude sections of several successive vowels occurring in a sample of speech one can, therefore, determine more easily and more exactly than from

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the degree of blackness on the spectrogram the conditions of amplitude, that is, stress, that prevail in the sample.

It is not wholly impossible for the viewer to recognize resonance bands also on a narrow filter spectrogram. If you look once more at Fig. 21, which was meant to illustrate by means of a narrow filter spectrogram the harmonics of a tune that glides from one frequency to another, you will notice also certain areas of greater blackness running from left to right through the entire spectrogram. These represent the formants of an [i] which the singer articulated throughout the melodic line. To produce *Fig. 29*, on the other hand, instead of sustaining one vowel at various pitches, the singer articulated the sequence [i - a - u] at an evenly sustained pitch. The resolution of the harmonics (the tenth harmonic is traced in white) therefore shows itself in parallel lines (cf. the sustained cello tone of Fig. 19). But now the three parts of the spectrogram representing severally [i - a - u] show three arrangements of areas of greater blackness, corresponding to the formants of the three vowels.

Fig. 30 illustrates four phonations of the vowel [i] which are distinguished from one another by intonation. The four intonation patterns employed are the four tones of Mandarin Chinese: even, rising, falling-rising, falling. So intoned, *i* has four distinct meanings: 'cure', 'suspect', 'chair', and 'easy'. As one would expect, the formants for the [i], so far as they are distinguishable, remain the same throughout the four words, while the harmonics (the tenth is again traced in white) follow the intonation, that is, the four tones of Mandarin Chinese.

Even though narrow filter spectrograms may, then, furnish some indication of formant position, it is nonetheless clear that, if our primary purpose is to show the resonance bands in order to distinguish with good visibility typical resonance bar patterns, that is, speech sounds rather than pitch variations, a broad filter spectrogram will serve us better.

In addition to the device which in the spectrograph regulates, if necessary or desirable, the sensitivity to powers and selects

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

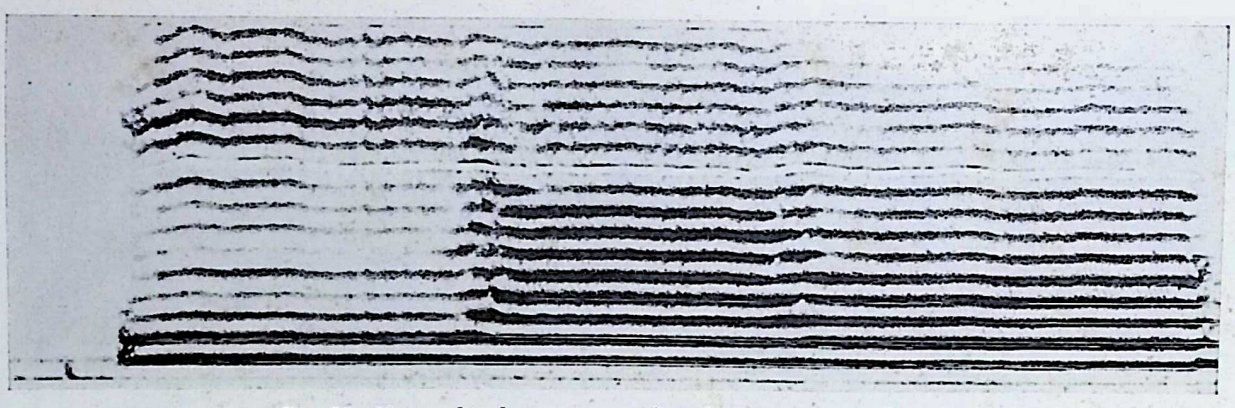


Fig. 29. Narrow band spectrogram, [i-a-u] sung at sustained pitch.

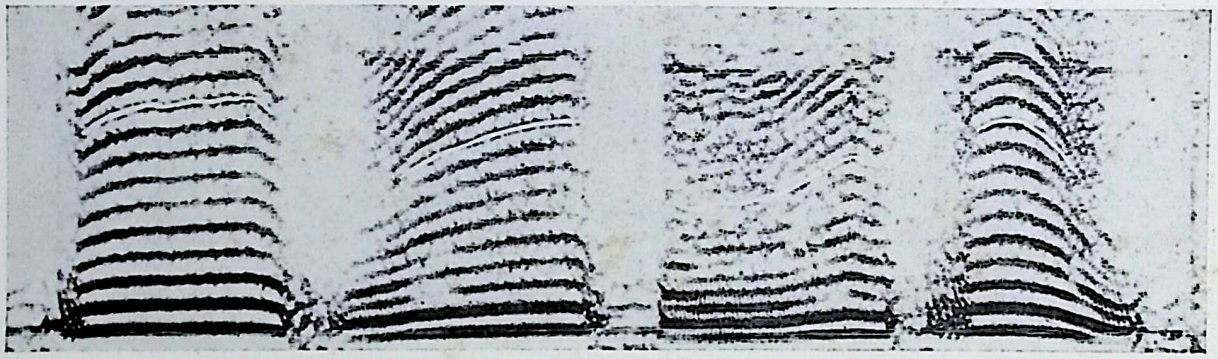


Fig. 30. Narrow band spectrograms, [i] articulated with four different intonation patterns (tones of Mandarin Chinese).

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

those above a chosen minimum which the experimenter wishes to be marked by the stylus, it is also possible to amplify or decrease the loudness of the speech sample that is recorded on the tape, before it enters the analyzing part proper of the machine, so that a low decibel recording can be strengthened and a high one tempered. By balancing these two adjustments of power, the one in the machine itself and the other of the recording to be analyzed, the spectrographer can, through a process of trial and error, or by means of calculation, arrive at the production of such a spectrogram as contains exactly those required contrasts and degrees of blackness which best visualize the significant components of the sound under examination. In fact, the commercially obtainable machine is conveniently so calibrated that generally acceptable results are obtained when the needle of the volume indicator hovers around 'zero'.

It appears, then, that for the purposes of a speech analysis which aims at visual results most distinctive in terms of segmental phonetics or phonemics, the best arrangement of filtering is the one employed to produce Fig. 28, because it reproduces best the significant, relevant features of linguistic sounds, classified as phonemes, at least of vowels. (Concerning non-vowel phonemes see below.) It follows from the preceding discussion that the several spectrographic representations fulfill specific tasks in the visualization of speech. (1) The broad band spectrogram displays the formants, that is, the typical patterns of power among the frequencies present in the course of an utterance. (2) The narrow band spectrogram displays the harmonics, that is, the melodic or intonational pattern of an utterance. (3) The amplitude section displays the spectrum, that is, the power, or the amount of stress, inherent in the component frequencies of a sound at a given moment. In other words, (1) shows what phonemes occur in an utterance, (2) displays the intonation of an utterance, and (3) indicates the stress pattern of an utterance. Of course, all three may be examined

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

not only separately but also in combination or juxtaposition (cf. Figs. 19, 28, 31).

Sounds other than speech sounds, such as musical tones, noises, the singing of birds, the humming of motors, the beating of hammers, all the sounds that surround us, require different types of spectrographic analysis, with different filters, different power passes, and often registrations that go beyond 3600~. The best installation for each task must be determined by experimentation, which will show just what feature in the sound is its most typical property and needs to be visualized in preference to others. But it is true that the *complete* spectrographic description of any sound, including phonemes, requires several kinds of spectrograms viewed together; for each spectrogram is capable of purveying some but not all information on the sound that is being examined. Needless to say, then, the spectrograph's usefulness is not exhausted by linguistic research, and surely not by providing the kind of spectrogram of speech which, because of its excellent and distinctive visual properties, has received the name of 'visible speech'.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XIX

SPECTROPHONEMES (VOWELS)

I have stated before that acoustically non-same, but similar, sounds are by each language categorized into classes of linguistically same sounds which are called phonemes. (Some linguists would demur on my requirement of acoustic, or phonetic, similarity. I have already explained, from several points of view, why I choose not to ignore the physical aspects in phonemic analysis.) It follows that to this linguistic sameness of classes of perception within the inevitable acoustic non-sameness of events of physical reality, there must correspond a linguistic sameness of classes of articulation within the acoustic non-sameness of events of articulatory reality. Phrasing this in other terms, I can say that no two speech sounds, or articulations, can be acoustically alike, because the physical act of articulation cannot be exactly repeated either by the same speaker or by different speakers; however, all members of a linguistic community speaking the same language have agreed as a matter of custom to recognize, and are able to imitate, sameness of sounds within fixed classes and within fixed scales of tolerance, so that all those phones, that is, events of articulation, which are linguistically identifiable with one another though not acoustically identical with one another, belong to the same class, which we call phoneme. In other words, we do not pronounce phonemes but phones; but we recognize, classify, and produce each phone within the limits of its class, limits which we learn to observe as relevant in the structured system of any language.

No two speakers posess exactly the same speech organs in terms of anatomy and physiology. Yet all speakers can learn to arrange their speech organs in such a way that, after allowance

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

has been made for personal peculiarities, linguistically same sounds, classifiable severally in phonemes, can be produced by all.

Lest I be misunderstood I want to make it clear that under sameness of class coupled with a requirement of phonetic similarity among all members of the class I do not have in mind an absolute rigidity of boundaries. If such rigidity and equilibrium were the rule, by the way, languages would not undergo changes in their phonemic structure—and we know that they do.

It may happen that in a language which has /ɛ/ and /æ/ as neighboring (in terms of place of articulation) palatal phonemes, the phonations [ɛ$^{\vee}$] and [æ$^{\wedge}$] occur either in some positions (positional variants) or habitually in the mouths of some speakers (idiolectal, social variants). ($^{\vee}$ and $^{\wedge}$ mean lowered and raised articulation, respectively.) This may reach the extreme point where a very low [ɛ$^{\vee}$] and a very high [æ$^{\wedge}$] are in fact phonetically identical; let us call this allophone [x]. Since the language imposes the necessity to observe phonemic distinctions, the following arrangements of phonations are possible:

Phonemic statement		/ɛ/	:	/æ/
Phonetic implementation	(1)	[ɛ]	:	[æ]
,, ,,	(2)	[ɛ]	:	[æ$^{\wedge}$]
,, ,,	(3)	[ɛ]	:	[x]
,, ,,	(4)	[ɛ$^{\vee}$]	:	[æ]
,, ,,	(5)	[x]	:	[æ]

Even when the boundaries of /ɛ/ and /æ/ are shifted, the phonemic distinction remains preserved. Hence in terms of articulation, phonetics, and spectrophonetics, it will be possible for the receiver to distinguish *bet* from *bat* at all times as soon as he has learned to recognize the scale of distinctions habitually employed by the user. This will not necessitate an explicit statement by the user, or his pronouncing *bet* and *bat* in order to establish the scale in the hearer's memory. For even though the speaker may be accustomed to saying and hearing

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

[bxt] for *bet*, the heard phrase "The pitcher was at [bxt]" calls unmistakably for the meaning *bat*. (The phrase "He raised his [bxt]" may still remain ambiguous. But no utterance, in normal conversation, is spoken in isolation, without a linguistic and situational context. Still, where all fails, ambiguity may remain until synonyms or rephrasings bring clarity. Such ambiguity is, indeed, the essence of a pun.)

But there is yet another possibility, namely,

Phonetic implementation (6) [x] : [x]

As long as this is a personal, occasional idiosyncrasy or oddity, it is dealt with by the hearer or the viewer of a spectrogram exactly as are the other cases which might give rise to misunderstanding or confusion. But if it becomes the norm in an entire speech community, then we are faced with a neutralization of phonemic contrasts, with a phonemic merger. Viewed diachronically, this is now a phenomenon of historical linguistics, which does not concern us here; viewed synchronically, we are simply dealing with a different phonemic structure in its own right, with a different set of phonemes and contrasts, which we need not and must not describe with reference to any other preceding or contemporaneous structure.

I admitted a concept of scale which gives a certain latitude of phonemic realization so long as phonemic contrast is maintained. Would it therefore be possible that in a language which has, like English of Tab. I, five palatal vowel phonemes plus /a/, a certain speaker may realize the /i/ by means of a phone as low as lies other people's realization of the /ε/ phoneme, provided only he squeezes between this [ε], which is as high as he chooses ever to go, and the [a], which is as low as anyone can go, the requisite four contrasts that remain? This would in fact mean that one and the same phoneme /i/ of the language can be implemented by [i], as done by the majority, and by [ε], as done by a minority, for whatever reasons. And if that is so, is it possible to stretch the concept of phonetic similarity among allophones of one phoneme so far as to include what are, by any measurement of speech performance, items so dis-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

similar as [i] and [ε]? (It may be objected here that the speaker of a language which has, say, only /i/, /a/, and /u/, may indeed be unable to hear or reproduce the contrast /i/ : /ε/ of another language, and that to him, therefore, [i] and [ε] are indeed similar. But this is a problem of comparative phonemics, a tricky subject that has been paid little attention in the literature so far, and it does not imply that the man uses, when he speaks his own language, a phone so far removed as [ε] to implement /i/. A Spaniard, for example, who has a scale of contrasts /i/ : /e/ : /a/, may be unable to hear or to reproduce the contrast between /i/ and /I/ when he hears or speaks English. Yet he will never implement his Spanish /i/ by means of what one might, on the basis of the Spanish scale, assume to be a sufficiently similar [I]. And anyone who uses [I] in Spanish speaks with a foreign accent. A Frenchman, curiously enough, may have great difficulty with the German /χ/ as in *Nacht*, although he has an allophone of uvular /r/ which phonetically is the equivalent of the normal [χ] allophones of the German /χ/: he pronounces *soir*, phonemically /swar/, as [swaχ].) Now if, in English, the kind of implementation of /i/ by means of either [i] or [ε] and the squeezing in of four contrasts between /ε/ and /a/ actually occurred with any kind of regularity, it would have the consequence that on the spectrogram, which is produced by a machine that has no culture and no language, hence no phonemes and no set number of phonemic distinctions, the visual near-sameness or overlapping of notations that allegedly represent distinct, non-same phonemes, would cause an illegible and insoluble puzzle. This kind of spectrographic chaos does *not* occur. The reason for this absence of spectrographic confusion must be that acoustic tolerance in speech does not go quite so far, and that speakers will on the whole refrain from putting each other to such severe acoustic tests of comprehension. I doubt, then, that any language puts upon speakers and hearers the burden of actually requiring them to make four significant distinctions within the small articulatory and acoustic (hence spectrographic) space

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

that lies between [ε] and [a]. Or, to place the responsibility where it belongs, on the users rather than on the language, human beings can hardly tolerate for long a phonemic arrangement of this type, any more than they have relished music whose significant intervals are narrower than a quarter-tone, or even a half-tone, of the conventional seven-tone musical scale. An overly demanding conventional, cultural segmentation of a physical continuum taxes too much our human physical limits of perception, and makes no allowance and leaves no room for human imperfections of hearing and performing. (To my knowledge speech intelligibility has not yet been scientifically and quantitatively investigated from this point of view.) True, phonetic or phonemic arguments based on the 'difficulty' of a sound are notoriously suspect because they are often relative and culture bound: the Czech trilled fricative /ř/, spelled *ř*, is difficult to utter for almost everyone but a native speaker of Czech; to him it is no more difficult than the trilled non-fricative /r/. Yet since in speech we are dealing with physical events which are produced and must be perceived by human organs of certain humanly inescapable limitations in terms of both motor skill and perceptual capacity, reasoning that involves impossibility and difficulty and ease of enunciation and hearing must not be rejected *a priori*—though they need by all means to be carefully scrutinized before being accepted.

It seems to me, then, that there is enough acoustically fixable sameness in the vast majority of implementations of one phoneme, and sufficient reluctance, for practical reasons, to convert the elasticity of phoneme boundaries into phonetic arbitrariness and licence, so that circumscription and acoustic definition of phonemes on an absolute scale, albeit one with elastic boundaries, are both feasible and necessary.

But I need not have recourse to the argument of practicality and the speakers' reasonableness. Recent phonemic theory offers a view of the phoneme which incorporates proper regard for physical reality and reestablishes order in the threatening

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

disorder of unlimited relativity. It provides that each phoneme is defined by a whole bundle of concurrent distinctive features. Up to this point in my discussion, when characterizing the phonemes /i/ /e/ /ε/ /a/, I spoke only of such oppositions as high *vs.* low, closed *vs.* open, front *vs.* center, all of which refer to articulation, in particular to the position of the tongue and the lower jaw. But I neglected such other important feature oppositions of acoustic nature as compact *vs.* diffuse, tense *vs.* lax, strident *vs.* mellow, grave *vs.* acute, etc. (These adjectives are used metaphorically, and are, not always felicitously, I think, chosen with an appeal to the synaesthetic sense of the user.) They appear spectrographically as the following oppositions, respectively: higher concentration of power in a narrow central region together with an increase of total amount of power *vs.* lower concentration of power together with a decrease in total amount of power; higher total amount together with a greater spread of energy in the spectrum and in time *vs.* lower total amount with lesser spread; higher intensity noise *vs.* lower intensity noise; concentration of power in lower frequencies of spectrum *vs.* concentration in higher frequencies; etc.

It is clear that the more a relation among a number of items is defined by conditions of opposition that keep the items apart, and the greater the number of oppositions per pair of items, the more accurately each item in these relationships becomes defined by absolute properties.

Let me illustrate this by reference to a game called 'Animal, Vegetable, Mineral'. For the purpose of identifying a certain subject on which several persons have agreed, the questioner asks, as skillfully as he can, a series of questions, but only such as may elicit a yes-no answer. He tries, in other words, to establish a number of binary oppositions, and each yes or no answer excludes all the members of one or the other class that is characterized by belonging to one or the other side of each binary opposition. Gradually the subject, through constant reduction in the membership of the pairs of classes the questioner opposes, becomes pinpointed. Of course, good guessing

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

is helpful—but on the other hand, bad guessing is harmful. Assuming that the subject was Shakespeare and that the questioner was informed that it was 'Animal', his binary opposition questions may take the following course: Living (as opposed to dead)? No. Human (as opposed to non-human)? Yes. Male (as opposed to female)? Yes. American (as opposed to non-American)? No. Asian (as opposed to non-Asian)? No. European (as opposed to non-European)? Yes. English (as opposed to non-English)? Yes. After 1700 (as opposed to before 1700)? No. At this point the questioner knows that he must further identify an Englishman who lived before 1700.

Now we may imagine that the human brain, when it receives a speech sound, goes through a similar series of yes-and-no questions that concern what we call distinctive features. Is the sound vocalic (as opposed to non-vocalic)? Compact (or diffuse)? Tense (or lax)? Strident (or mellow)? And so forth. The more such binary opposition questions are asked and answered, the closer the hearer comes to identifying the specific item. I need hardly mention the unimaginable speed with which the brain operates, the enormous number of such yes-no questions it can sort in a brief time, and how rapidly each auditory impression of an articulation can therefore be phonemically classified.

In addition, the knowledge that speech, and presumably a meaningful message, is perceived, helps a good deal in that it permits some guessing. Furthermore, languages are redundant on all levels, phonemic, morphemic, and syntactic (excepting an ideal, constructed language, which is yet to be created, or a logical super-'language' like mathematics). This fact, by the way, seems to be often overlooked by those linguists who operate with all linguistic signals on all levels as if all of them were in all instances of equal importance, and as if none of them were at any moment expendable. The redundancy and the expendability of some signal occurring in a context, as it generally does in communication, lighten the burden imposed on speakers and hearers. Thanks to this over-signaling we may

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

understand, and be understood, even if for some reason (noise interference, low-fidelity electronic circuit, inadequacy of articulation or perception) the acoustic signals do not in all details and on all levels correspond fully to those which an analysis of the language in question has established. It follows from this that linguistic experiments concerned with intelligibility, whether they are testing the fidelity of an electronic circuit or the proficiency of a student in comprehending a foreign language, must not overlook the aid which a context affords the listener in identifying the signal which the experiment is testing. Phonemic tests of substitution, therefore, which operate with minimal pairs (*tack - tag*, *tag - tug*) outside of a context, make in fact more rigorous demands upon the listener or his brain than are generally encountered in normal communication, which partakes of both a linguistic and a situational context. But, of course, there is always a limit of tolerance, or a minimally adequate requirement of performance. The context and redundancy in linguistic communication, then, absolve us from the necessiy of articulating perfectly, and perceiving perfectly, model signals. Indeed it is more than questionable whether an artificially constructed vocal code which has no redundancies would serve us well. It would necessitate strict adherence in detail to the model signals in articulation, acoustic sterilization of the conducting medium (no noises in the air beside the signals transmitted, electronic circuit of natural fidelity), and perfect hearing and concentration on the part of the receiver. These conditions are virtually impossible of fulfilment in this world among humans.

Whether we like it or not, phonemes are not classes of *things* with mathematically or biologically or physically fixed boundaries. They are, rather, classes of behavior which are circumscribed and defined not by laws of nature, but by somewhat elastic and historically unstable rules of conduct, upheld by a desire and a necessity for conformity and uniformity, however ephemeral. All such classes of behavior, linguistic or other, constitute the cultural and social part of ourselves and of our

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

environment. Change is of their essence, and time alone a sufficient agent to provoke it—although many other factors may be operative. Every language is a structured system of such classes of behavior and must be described and understood as such. The part of linguistics which deals with language as a social system belongs therefore in the behavioral sciences. But in regard to the machinery and the material of the production, transmission, and the perception of language, linguistics partakes also of the natural sciences. A true description of language must draw its information from both. Hence spectrography, or any other experimental technique of linguistic analysis, constitutes a particularly felicitous and fruitful collaboration insofar as the spectrograph provides an objective, acoustically ordered quantitative analysis, the natural science part, and the spectrographer the subjective, culturally ordered, qualitative analysis, the behavioral science part. This presupposes, of course, that the linguist-spectrographer should have mastered both branches of linguistics, the natural and the cultural science one. Hence every linguist, even though he may not aim to be a spectrographer, ought to be at least acquainted with the principles and results of spectrographic analysis of language, just as the spectrographer, even if language is not his chief interest, ought to be conversant with the rudiments and basic tenets of linguistics.

Having noted the peculiar socio-cultural, in our case linguistic, sameness of sounds on the one hand, and on the other hand the ability of the user of the language to use and understand such linguistic sames despite the natural obstacles to acoustic sameness, we naturally expect and require that a good visual representation of speech sounds render recognizable linguistic sames despite the natural acoustic variety of events, and despite the cultural neutrality and insensibility of the spectrograph. In other words, we expect and require that each phoneme of a language have a typical and recognizable spectrogram, although we are now reconciled to the fact that of necessity the limits of each phoneme cannot be set with mathematic precision and irrevocability.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

It is obvious that the formants and resonance bands must play as important a part spectrographically as they do acoustically in the composition of vowels, and that therefore vowels should be typically distinguishable from one another by their arrangements of formants. It follows, therefore, that the spectrograms and amplitude sections of the vowels as pictured in Fig. 28 are representative not only of the speech events here visualized but also of the classes, that is, the phonemes, to which they severally belong. This is not just an assumption but a fact: every phonation of any one of the vowels shown resembles recognizably and typically every other phonation spectrographically as well as acoustically. Of course, since the absolute position of the formants on the frequency scale and their absolute power on the power scale are determined by phonetic and anatomic circumstances which vary from speaker to speaker and from phonation to phonation, since they are, in other words, non-phonemic and non-linguistic, one must argue that it is, apart from their occurrence, their relative position to one another which determines their linguistic value. Hence the spectra and amplitude sections of Fig. 28 are not only spectrophonically true, but also spectrophonemically valid—in the sense that articulatory and acoustic classification of sames into phoneme classes corresponds to visual classification of sames (configuration of formant bands) into SPECTROPHONEME classes.

If there did not exist such classes of spectrographic notations of speech sounds which I have just called spectrophonemes, then one of the earliest applications of spectrography, visible speech, which is based on the recurrence of such classes, would be impossible. It would also be impossible to read phonemes off a spectrogram, which a spectrographer can do. Of course, this presupposes that the reader not only can decipher the secondary visual code, the spectrogram, but that he has already mastered the primary code, the language he is reading. I have not neglected to say earlier that the spectrograph alone makes no decision as to what a phoneme, or a spectrophoneme is. Just as a phoneme can be defined only with reference to the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

larger frame, which is the language of which it is a structural feature, with physical and acoustic (phonetic) considerations playing an important controlling part in that they prevent the linguist from putting phonetically wholly unrelated items into one class, so the spectrophoneme is determined also by a preceding decision of the linguist on the number and types of phonemes in the language he is analyzing, with physical and acoustic (now spectrophonetic) considerations again playing an important controlling part in that they prevent the linguist from putting spectrographically wholly unrelated items into one class.

Phonetic details may escape the ear of the analyst, especially in a foreign language upon which he is likely to impose, depending on the degree of his skill, many or few of the prejudices of his native phonemic scale. Hence the spectrogram is useful in showing these details visually, since the machine has no distorting prejudices. Whatever shows on the spectrogram is, therefore, an acoustic reality, regardless of whether the analyst 'hears' it or not. And only those acoustic realities which are not spectrophonemically unrelated may be assigned to the same phoneme. It is in this manner, then, that spectrography, and the concept of the spectrophoneme, are of aid in phonemization, in the phonemic analysis of a language. But, to repeat, the spectrogram of and by itself gives no information on either phonemes or spectrophonemes.

But since human shortcomings, especially as regards cultural prejudices, what one might call phenomena of cultureblindness, are ubiquitous, they will beset, if not the spectrographic record, then at least the reading of it. For just as the hearer may not always 'hear' the full acoustic truth, so the viewer of the spectrogram may have acquired, thanks to protracted experiments with primarily one (probably his native) language or several languages, certain optical prejudices which prevent him from 'seeing' the full optical (and by implication acoustic) truth. Hence the spectrographer, too, must as far as humanly possible divest himself of whatever cultureblindness he may have fallen

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

prey to. The fact that the spectrogram itself represents an objective reality does not protect its viewer from mis-seeing it; for the registration of reality by an unprejudiced machine and its interpretation by a human observer are two very different performances. But this may be said in favor of the spectrographic over the audible record, and this is, in fact, one of the principal reasons for creating a spectrograph: the spectrographic record is permanent, it can be examined and measured at leisure, it can be experimentally varied and modified to emphasize now one now another acoustic peculiarity; it is, in short, analytic and quantifiable in its significant aspects. All this the audible evidence alone (even though it may now be made permanent and repeatable thanks to recording devices) is not.

Now if there are spectrophonemes that are acoustically localizable, there must also be phonemes that are articulatorily localizable, for spectrophonemic sames do parallel acoustic, hence articulatory, sames (excepting those rare cases where compensatory articulation takes the place of normal articulation, such as occurs in physically damaged speakers). And this refutes, to my mind, the view, stated more than twenty years ago and repeated ever since in one form or another, that "the presence of 'phoneme-features' as positive, additive entities in the sound waves [N.B.—this refers to oscillograms not spectrograms] is not demonstrable, and there is no reason to believe that it will be". It is now, in fact, so demonstrable. On the same grounds I also reject the opinion that a phoneme cannot be acoustically described. I say it can—with the proviso, just stated, that its definition include proper regard for its being also a class of conventional, cultural behavior, although its material is of the physical world.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XX

SPECTROPHONEMES (NON-VOWELS)

Since the vowels give the spectrographically most satisfying and distinctive pictures, or spectrophonemes, and since my goal has been chiefly to illuminate the principles of spectrography rather than to train expert spectrographers, the non-vocalic phonemes have been rather neglected in the preceding pages of this book. I shall now remedy this lack at least to the extent which is germane to my present purpose.

All voiced phonemes, like the vowels which are the voiced phonemes *par excellence*, are produced in part by vibrations of the vocal lips. On the spectrogram, therefore, voiced consonants are recognizable by the presence of voice as shown in the glottal band, at least. But since the vibrations of the vocal lips for consonants are weaker and often of shorter duration than for vowels, the registration at the bottom of the spectrogram representing the fundamental glottal pitch is correspondingly less dark and often shorter than the corresponding registration in the vowels; and the formants F_1, F_2, F_3, and higher, are even less in evidence. But lack of darkness can be corrected, if desired, by increasing the volume of the recording, or the sensitivity to power of the spectrograph, or both, so that the presence of voice can be registered more clearly. In any event, for purposes of reading a spectrogram of speech a mere indication of the presence of voicing is quite sufficient, because the precise location and strength of the glottal resonance band, being equivalent to glottal pitch and its loudness, are linguistically not relevant. The unvoiced consonants, however, by definition lack the bottom resonance band, no matter how the spectrograph is adjusted.

Some of the voiced consonants, as we have seen, share with

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

the vowels the property of being resonant, and they exhibit, therefore, also the feature of resonance bands. To this group belong English /m/ /n/ /ŋ/ /l/ /r/. The location and shape of their resonance bands is determined to some extent by the nature of the following or preceding vowel, just as, in turn, the beginning and the end of the resonance bands of a vowel is in some measure shaped by the character of the preceding or following consonant.

The next important distinction among the consonants is whether they are plosives (stops) or continuants.

The stops are created generally by a sudden release of a stricture or of a closure, and depending on the place of the explosion one distinguishes labial, dental, palatal (at different spots along the palate, hence pre-palatal, medio-palatal, and post-palatal or velar varieties), laryngeal, and glottal stops. Since the duration of these sounds (except for the period of silence during which pressure is built up and which can be extended) is naturally very brief (how long does the release of a closure last?), there is small chance of their leaving a distinctive and typical picture of themselves on the spectrogram. All of them, however, are characterized by a period of silence which immediately precedes the release and during which pressure for the explosion is gathered behind the obstacle. On the spectrogram such silence appears, of course, as absence of power, that is, as a blank area without stylus traces. But voiced stops may show the bar of glottal voicing throughout the articulation, in both the silence and the release: the oscillation of the vocal lips sets in before the release. (This can be easily verified by a simple experiment. By shielding one's ears so as to exclude external noises and then pronouncing a [b], [d], or [g] in isolation, one may observe a humming inside one's hand preliminary to the release of the explosion, which is caused by the activity of the vocal lips. Slow-motion films show a definite period of increasing frequency and amplitude of vibration of the vocal lips extending from their original position of rest to the attainment of the oscillation suitable for the tone that is

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

produced. Singers must be carefully taught to shorten and weaken this gliding toward the proper note.)

Such blank spots on the spectrogram, with or without a glottal resonance bar, can, of course, deliver no distinguishable pictures of the various plosive phonemes of a language. It has been discovered, however, that the different types of plosives leave severally distinctive marks of their articulatory peculiarity in the off-glide of the preceding and the on-glide of the following phoneme and, by implication, spectrophoneme. Especially if either of these is a vowel, its end or beginning shows on the spectrogram a finite number of typical variations and changes in shape and direction, called formant transitions, each of which affords a clue concerning the class to which the following or preceding plosive belongs. (But note that the spectrophonemes of the vowels of Fig. 28 have been so cut as to remove the formant transitions, thus affording a view of the 'pure' vowels.)

In fact, all neighboring sounds color one another. A very strong influence of a sound upon its neighbor may cause partial or complete assimilation of the two, a phenomenon of great importance in historical linguistics. Partial assimilation causes the unvoiced /s/ of *observe* to sound as a voiced [z] because of the continuing of the voicing from the /b/; complete assimilation in the same word, although exerting itself in the opposite direction, has fully converted the voiced /b/ of Latin *obseruare* into the unvoiced /s/ of Italian *osservare*.

Once the explosion has taken place there follow certain noises caused mainly by the friction along whatever obstacles or strictures the escaping air passes on its way out of the mouth. But these frictions, being mere noises and not tones, cannot be oscillographically represented by sine waves. Hence they do not appear spectrographically in any easily identifiable or typical form, least of all as formant designs, but only as random vertical striations on the paper, thus indicating power at random frequencies. There can be no question of fundamental and overtones, but only of an odd composition of fre-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

quencies, which give the acoustic impression of scratching, popping, clicking, grating, and the like.

Non-plosive, continuant sounds of this type may in any language be not only noises accompanying the plosive phonemes, but whole phonemes. The fricatives, nasals, laterals, trills, and spirants, produced at various points in the pharynx, mouth, and nose belong in this class. (Some of them, being resonants, show also resonance bars, as I mentioned earlier.) Representing phonemes, that is, classes of speech sounds, on the spectrogram, they too may be expected to be susceptible of classification in spectrographic types, or spectrophonemes. But learning to recognize them, especially the non-resonants, is considerably more difficult than was the recognition of the spectrograms of vowels. For these continuant phonemes made up chiefly of noise rather than tone, while they last long enough to be visible and segmentable within the stream of speech, are by nature capricious and do not cluster as closely around a spectrographic norm as do the vocalic phonemes. Moreover, the notation which they leave on paper is not as distinct and striking as one would wish. If we analyze only up to 3600~, as we generally do, it may well happen that we omit thereby important and significant power components of these noises, a risk scarcely attendant upon limiting spectrograms of vowels at this frequency.

The spectrograph has yielded confirmation of an important tenet of phonetics, long held but now spectrographically attested in the formant transitions, which by some is couched in the phrase that all speech sounds are glides. The term 'glide' is employed here somewhat loosely, and is merely meant to indicate that, in the stream of speech, no sound is an island, a completely independent segmentable unit throughout its articulation. Phonetically the name glide has its source in an historical or comparative view of language and refers generally to the addition of a transitional sound that connects two phonemes. Synchronically, a speech event is what it is and as it is, and only in diachronic terms can we say that something was

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

'added' in the pronunciation of a sequence of phonemes, some extra sound generally due to inertia of the vocal organs and for the purpose of easing difficult (in the terms of difficulty commented on above) clusters. Such an added glide may or may not eventually attain phonemic or graphemic status.

For example, the /b/ in French *chambre* 'room' is, historically seen, a glide which fills the transition from /m/ to /r/ of a late Latin *camra,* earlier *camera,* whence the posttonic unaccented /e/ had been dropped (syncopated). The opening of the lips after the voiced /m/ in preparation for the voiced /r/ automatically makes the noise of a voiced bilabial plosive, which, it so happens in Latin and French, is a phoneme of these languages generally represented by the spelling *b*. Sometimes, though not in the *cam(e)ra* > *chambre* example, or in Latin *hum(i)lem* > French *humble,* English *humble,* the conservatism of orthography prevents such a glide from finding its way into spelling. We continue to spell German *fünf* although most of the time we pronounce it more like *fümpf* (with assimilation of the dental /n/ to the labiodental /f/, so that the insertion of the glide /p/ produces a wholly labial cluster), or spell English *family* but pronounce it *fambly.* The pronunciations *fümpf* and *fambly* are by many speakers considered sub-standard, partly no doubt because of the conservative spelling of the standard literary language. *Humble* and *chambre,* however, are standard in pronunciation and spelling (there being no **humle* or **chamre*), partly no doubt because the *b*-less etyma are long since forgotten and cannot at this late date be exhumed and resuscitated with any hope of success by even the most history-conscious speaker.

When someone says that all sounds are glides, he does not mean this in the technical historical, diachronic sense just stated. He means, rather, that the articulation of all sounds is in some measure influenced by their phonetic surroundings, especially at the areas of contiguity. The spectrogram does indeed show that the full duration of a phoneme is not an even, homogeneous articulation. Instead, it consists of an on-glide and

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

an off-glide of considerable length, both of which are acoustically and spectrographically shaped by the preceding and following phoneme, respectively, enclosing but a brief core which alone represents, so to speak, the 'pure' phoneme and which alone contains *all* the characteristic spectrographic details and *all* the standard acoustic distinctive features without distortion and interference.

The total length, then, of a unit which on the spectrogram can be sectioned off as the realization of a phoneme, is to a considerable extent occupied by an articulatory movement toward the core of the phoneme, and an articulatory movement away from it toward the next phoneme. I have mentioned before that characteristic recurrent features of such on-glides and off-glides, the formant transitions, provide in some instances, especially in regard to plosives, the most important and often unique indication on the nature of phonemes not otherwise clearly and unambiguously registered by the spectrograph. It follows that in the normal stream of speech we do not insist or dwell long on a phoneme, but that acoustically, as well as spectrographically, its appearance of duration often includes movements towards and away from it. And in this sense every speech sound is indeed a glide.

This becomes fully confirmed when one has an opportunity to follow visually the processes of articulation in the speaker's vocal organs. Direct observation and moving pictures show beyond any doubt that articulations are not segmented but that they succeed one another in an extraordinarily varied and complex series of movements. One realizes how difficult and involved a process speaking really is. Man is the only animal on earth possessing the requisite nervous and cerebral equipment to cope with this task both physically and intellectually.

On the other hand, there is no doubt that it is possible to find the discrete units which make up this apparently integrated continuum. Even if modern linguistics did not occupy itself with this problem in its phonetic, phonemic, and acoustic aspects, the very existence of alphabetic writing, invented with-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

out the aid of modern professional linguists, testifies to this peculiar analyzing and classifying faculty of man. We may perhaps not assume that the first alphabetic notation of a language, except those accomplished in our days by a trained linguistic technician, satisfied all the rigid requirements of modern phonemization (any more than do many current orthographies which bear the weight of an unyielding convention). Yet it is nonetheless true that the inventors of such ancient alphabetic scripts (and not all ancient, or modern, scripts are alphabetic) performed a segmentation of the stream of speech into discrete units, or phonemes. In fact, most people, including layman, will more often than not find themselves in agreement, apart from a relatively small number of problems that only the phonemicist can solve (and some not even he has solved, such as whether the initial sounds represented by the spelling *ch* in English *chin* are to be analyzed as one phoneme /č/ or as a cluster /tʃ/), on how to segement a continuum of speech in their own, and sometimes even in a foreign or in an unknown, language.

Another type of segmentation of the stream of speech, the syllabic, is also one on which most speakers of any language would readily agree; everyone knows how to divide an utterance into syllables. Yet an acoustic, linguistic, scientific definition of the syllable is still lacking, despite many attempts. Here, too, spectrography appears to be promising, especially the spectrographic representation of formant transitions. Experiments set up *ad hoc* have shown that a spectrogram of a sequence like *apa,* when it has formant transitions in the first but not the second /a/, represents an acoustic *ap-a,* whereas the presence of formant transitions in the second but not in the first /a/ renders a syllabification *a-pa.* In non-experimental, normal language spectrograms, however, these criteria have not yet been consistently observed and have exhibited individual variations in various speakers. In any event, clarification of this problem also will eventually be made easier through spectrography.

It would exceed the aim of this book if I attempted to in-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

struct the reader in the deciphering and the analysis of spectrograms, or to enumerate the many ways in which they are useful in the experimental study of language. Besides, a theoretical disquisition on these subjects, no matter how long and complex I should make it, would remain as incomplete as a course in chemistry without work in the laboratory. Spectrography, too, is best and most efficiently studied and taught in a linguistic laboratory where experimental work can be done under expert guidance.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

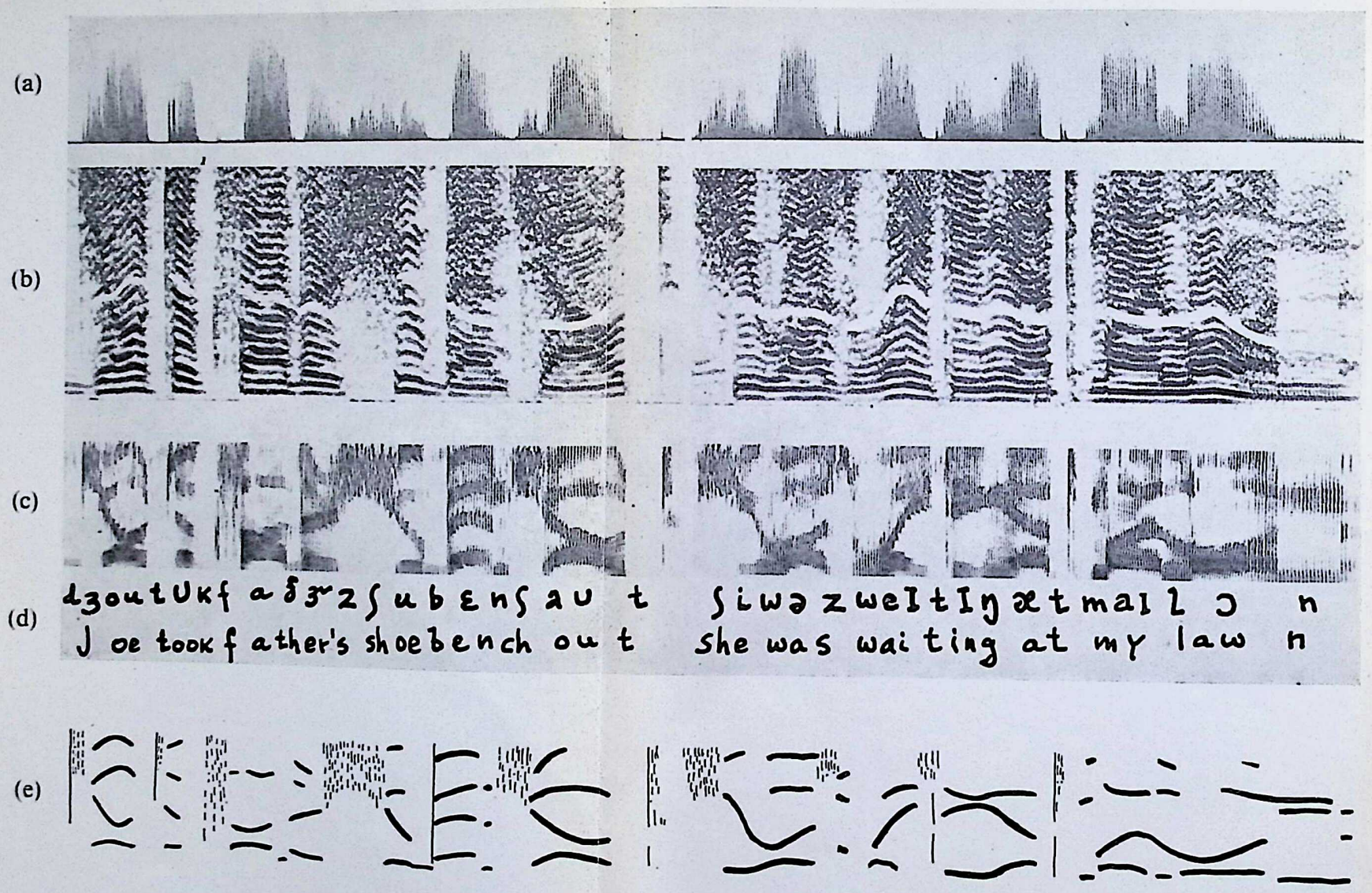


Fig. 31. Spectrograms and schematic.

(a) Continuous amplitude display; (b) narrow band spectrogram; (c) broad band spectrogram; (d) phonetic and alphabetic transcription; (e) schematic of (c).

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

XXI

SPECTROGRAPHEMES

Despite their complexity, speech sounds are acoustically recognizable, and spectrographically definable, on the basis of a fairly restricted number of factors, in particular by the relative location and shape of the resonance bands, and the less conspicuous but still characteristic distribution of some random frequencies.

This being so, it should be possible to design, by hand, spectrograms which trace out a schematic graph of phonemes, or spectrophonemes, omitting from the sketch all the details of negligible or no linguistic significance. Such attempts have been made and the designs consist mainly, as one would expect, of resonance bands, including such significant variations as formant transitions.

Fig. 31 represents various visualizations of the sentences: *Joe took father's shoebench out. She was waiting at my lawn.*

The top graph is a continuous amplitude display. But it is not of the spectrum kind, like those of Figs. 11, 13, 19, 27, 28, which display frequency on the abscissa and power on the ordinate; it is, rather, of the oscillographic kind, displaying time on the abscissa and power on the ordinate. The picture is equivalent to taking the bottom half of an oscilloscope pattern and turning it over onto the upper half of the pattern, and then filling in the area under the curves. Thus we have gained the parameter of time, but lost that of frequency.

The second graph is a 45~, or narrow band, spectogram; a double vertical scale was used for better separation of the harmonics, which are clearly visible (with the tenth traced in white).

The third graph is a 300~, or broad band, spectrogram, showing the location of the formants.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

The bottom graph is a schematic of the broad band spectrogram. Its main features are the contours of the formants, the most significant and distinctive spectrographic components of speech. Also some important striations (for fricatives) are drawn, and proper attention is paid to formant transitions.

Now if this design is in fact an adequate visualization of phonemes, then its transformation into sound should deliver an adequate reproduction of the same phonemes. The experiment was made by means of a so-called pattern playback. This is a spectrograph running in reverse, as it were, though of course equipped with an entirely different scanning and filtering arrangement that is based, like the tone band running alongside the picture band in a talking film, on optical scanning and on translation of optical impulses into sound. Into this machine was fed the hand-drawn pattern spectrogram, and from it issued an auditory signal. The results were, in fact, acoustically and linguistically adequate—but no more than that. The sounds emanating from the machine were not human speech (indeed other known synthetic speech machines, operating with a greater number of component parameters than three formants, give a closer approximation to natural language), yet the performance resembled speech sufficiently in its essentials so as to form an intelligible utterance.

Improvements of this operation are contingent only upon a perfection of the spectrographic pattern sketches, which are no more than hand-made recordings (similar to the perforated scrolls which are run through player pianos and activate the keyboard), and upon refinements in the playback machine. Actually, nothing but the accurate playback of a complete spectrogram, or better of several fused spectrograms, containing all, the redundant as well as the linguistically distinctive, elements and parameters of human speech sounds, can produce something that acoustically approaches real talk. Yet there is no doubt that the schematic spectrograms generating schematic speech are of enormous scientific importance for investigation and experimentation.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

But they are also of practical value. By means of them one can, for example, empirically determine the threshold of recognizability which electronic encoders and decoders of speech, like telephones, radios, amplifiers, etc., need in order to satisfy certain requirements of fidelity. Phrased differently, they allow us to determine how many redundant features we may omit without impairing intelligibility. Communication engineers are particularly concerned with this aspect of speech research. When they are required to design such devices, they keep in mind, among items of technical moment, the fact that accuracy of reproduction also bears upon the cost of manufacture and the market price to the consumer. By reducing fidelity to the minimum tolerable for the purpose of the machine, a much cheaper instrument may be fabricated; any superfluous increase in fidelity raises unnecessarily and steeply the price of the product and reduces its utility.

There are still other practical applications of spectrography which one may envisage, though their realization lies in the future. It is easily imaginable, for instance, that persons can be taught to write spectrographically, so that messages can be transmitted coded in a spectrographic instead of the conventional alphabet, in SPECTROGRAPHEMES rather than in conventional graphemes. One might even design spectrographic characters for a printing press, so that whole books could be printed in spectrograms. Perhaps this would not be a great gain over customary typography; but printing spectrographed messages on magnetic tape would. Such tapes could be fed into a spectrogram-playback, and a book could be listened to instead of read. It is, of course, even now possible to 'listen' to a book by means of tape and wire recordings, but no matter how many copies of an original tape or wire are quickly and cheaply fabricated, the initial matrix-recording has to be made by live voice. It is at this point that a spectrographic printed tape would save much time and effort, and money, even though the audible result might not equal a natural voice. But for a number of purposes, where only the content and not the

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

acoustic, including the aesthetic and literary, quality of the message is important, this would be no great drawback. Moreover, it is certain that even this shortcoming could be remedied if good acousticians and engineers set their minds to it.

Possibly the engineering difficulties and the time needed for the development of a marketable machine of this kind will turn out to be greater than I imagine. But the marketableness of a product and its price are a function of its usefulness. And I daresay that once communication engineers design a machine that is useful to many people, it will also attract buyers and encourage further improvements. Just like the phonograph and the wire and tape recorder, which not so long ago were mere gadgets and experimental contrivances, the spectrographic recorder and playback may become a household article.

A great stride forward in this domain was made by the recent development of the 48 channel spectrograph. All spectrograms described so far in this book are recordings of utterances lasting only 2.3 seconds each. To obtain a visualization of longer stretches of speech it is necessary to produce a whole series of spectrograms, and then mount them or paste them together in a row. The process is awkward and time-consuming since the production of each single spectrogram takes much longer than the analyzed utterance itself lasts. The 48 channel spectrograph, on the other hand, works like a movie camera and takes a continuous moving picture of the spectrogram. In this arrangement it is of course impossible to obtain spectrograms of the kind discussed so far, which are the result of two-hundred successive filterings of two-hundred playbacks of the entire recorded message, and the tracings of two-hundred successive stylus runs. This multiplicity of scannings was necessary, it will be remembered, so as to eliminate the cumbersome device of two-hundred or more filters in the machine. To allow the kind of photographing which the new machine does it is necessary, therefore, to fall back on the installation of multiple fixed filters which can operate simultaneously,

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

functioning somewhat like those of Fig. 14. But instead of two-hundred of them, the 48 channel continuous display spectrograph employs, as its name indicates, only forty-eight, which are scanned at the rate of five-hundred times per second. All filters up to a center frequency of 4000~ have a band width of 300~, those covering the higher frequencies have successively wider filters, and the first filter covers the frequencies between 0~ and 200~. For our present purposes it is superfluous to state the reasons for all this, or to discuss in detail the resultant spectrograms: the principles of analysis and display are the same in the older and the newer type of spectrograph, with each having certain advantages and disadvantages. (It might be noted that the 48 channel spectrograph can also be set to produce alternatively, or even simultaneously, a rapid succession of snapshots—64 frames per second—representing amplitude sections like those of Figs. 19 and 28. We have noted the inability of the ordinary spectrograph to accomplish this: in Fig. 31, we had to be content with an amplitude display of the oscillographic type. Now we may obtain something like a moving picture of amplitudes at certain frequencies also.) But as concerns the recognizability of formant patterns, that is phonemes, the 48 channel spectrograph is completely adequate. We are, then, in possession of a machine that delivers continuous display spectrograms (present cameras hold reels up to 30 minutes lenght) which, since they are photographed on film, are permanent records. These records can, like any film, be copied rapidly and accurately as often as one wishes, so that, in addition to their purposes for research, these spectrograms could also be employed for playback. They will, of course, render a much better approximation to human speech than the previously mentioned hand-drawn schematic spectrograms, since they represent speech more faithfully (as faithfully as a 48 channel spectrograph can). In fact, we are dealing here with a process of recording and playback similar to that of the sound track of a talking picture, except that the optical representation on the track of the spectrograph is spectrographic, hence lin-

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

guistically interesting and suitable for linguistic and acoustic experiments of a special type.

The use of the spectrograph, of whatever type, as an instrument for linguistic research and experimentation is of immense promise—and its exploitation has barely begun. In the long run, I daresay, this experimental rather than the practical aspect of its utility will be of greater interest and importance to the scholar. True, the spectrograph is not the only machine which a well-equipped linguistic laboratory must have, and it is not the one which fully analyzes *all* the signals and cues contained in speech. But, as things stand now, it is the most useful and versatile apparatus available, in particular for the linguist. The results obtained so far show already the enormous weight which the findings can bring to bear upon all branches of linguistic science, both descriptive and historic.

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

SELECTED BIBLIOGRAPHY

(For abbreviations see at end)

L. L. Beranek, *Acoustic measurements* (New York, 1949).

C. Cherry, *On human communication* (London, 1956).

C. Chiba and M. Kajiyama, *The vowel, its nature and structure* (Tokyo, 1941).

F. S. Cooper, "Some experiments on the perception of synthetic speech sounds," *JASA* 24.596–606 (1952).

P. Delattre, "The physiological interpretation of sound spectrograms," *PMLA* 66.864–875 (1951).

—, A. M. Liberman, and F. S. Cooper, "Acoustic loci and transitional cues for consonants," *JASA* 27.769–773 (1955).

—, —, and —, "Speech synthesis as a research technique," *PICL* VII (London, 1952).

—, —, and —, "Voyelles synthéthiques à deux formantes et voyelles cardinales," *MPh* 96.30–56 (1951).

—, —, —, and L. J. Gerstman, "An acoustical study of the acoustic determinants of vowel color," *Word* 8.195–210 (1952).

P. Diderichsen, "The importance of distribution versus other criteria in linguistic analysis," *PICL* VIII (Oslo, 1957).

H. W. Dudley, "Fundamentals of speech synthesis," *BTS* Monograph 2648 (1956).

H. K. Dunn, "The calculation of vowel resonances and an electrical vocal tract," *JASA* 22.740–753 (1950).

C. G. M. Fant, "Modern instruments and methods for acoustic studies of speech," *Royal* [*Swedish*] *Institute of Technology, Speech Transmission Laboratory*, Report No. 8 (1957).

—, "Speech communication research," *IVA* 24.331–337 (1957).

—, "Transmission properties of the vocal tract with application to the acoustic specification of phonemes," *Acoustics Laboratory, M.I.T.*, Techn. Report No. 12 (1952).

—, *Acoustic theory of speech production* (The Hague, in press).

E. Fischer-Jørgensen, "Acoustic analysis of stop consonants," *MiPh* 2.42–59 (1954).

—, "What can the new techniques of acoustic phonetics contribute to linguistics?," *PICL* VIII (Oslo, 1957).

J. L. Flanagan, "Automatic extraction of formant frequencies from continuous speech," *JASA* 28.110–117 (1956).

H. Fletcher, *Speech and hearing* (New York, 1953).

D. B. Fry, "Duration and intensity as physical correlates to stress," *JASA* 27.765–768 (1955).

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

—, "Perception and recognition in speech," *FRJ* 169–173.
M. C. Green, *The voice and its disorders* (New York, 1957).
M. Halle, "In defense of the number Two," *SJWh* 65–72.
—, "The strategy of phonemics," *Word* 10.197–209 (1954).
—, G. W. Hughes, and J.-P. A. Radley, "Acoustic properties of stop consonants," *JASA* 29.107–116 (1957).
E. Haugen, "The syllable in linguistic description," *FRJ* 213–221.
R.-M. S. Heffner, *General phonetics* (Madison, 1950).
C. F. Hockett, *A manual of phonology* (Baltimore, 1955).
G. W. Hughes and M. Halle, "Spectral properties of fricative consonants," *JASA* 28.303–310 (1956).
W. Jackson, ed., *Communication theory* (London, 1953).
R. Jakobson and M. Halle, *Fundamentals of language* (The Hague, 1956).
—, C. G. M. Fant, and M. Halle, "Preliminaries to speech analysis," *Acoustics Laboratory, M.I.T.*, Techn. Report No. 13 (1952).
L. G. Jones, "Preliminary phonetic segmentation," *SJWh* 117–119.
M. Joos, "Acoustic phonetics," *Language Monograph* No. 23 (1948).
—, ed., *Readings in linguistics* (Washington, 1957).
L. Kaiser, ed., *Manual of phonetics* (Amsterdam, 1957).
H. Kurath, "The binary interpretation of English vowels: a critique," *Language* 33.11–122 (1957).
P. Ladefoged, "The classification of vowels," *Lingua* 5.113 (1956).
—, and D. E. Broadbent, "Information conveyed by vowels," *JASA* 29.98–104 (1957).
I. Lehiste and G. E. Peterson, "Vowel amplitudes and linguistic stress in American English," *JASA* 31.1959 (in press).
D. Lewis, "Vocal resonance," *JASA* 8.91 (1936).
A. M. Liberman, "Some results of research on speech perception," *JASA* 29.117–123 (1957).
J. C. R. Licklider and G. A. Miller, "The perception of speech," *Handbook of experimental psychology*, ed. S. S. Stevens (New York, 1951), 1040–1074.
L. Lisker, "Linguistic segments, acoustic segments, and synthetic speech," *Language* 33.370–374 (1957).
J. Lotz, "The structure of human speech," *Transactions of the New York Academy of Sciences*, Ser. II, 16.373–384 (1954).
A. Malécot, "Acoustic cues for nasal consonants," *Language* 32.274–284 (1956).
P. Menzerath, and A. de Lacerda, *Koartikulation, Steuerung und Lautabgrenzung* (Berlin–Bonn, 1933).
W. Meyer-Eppler, "Die Spektralanalyse der Sprache," *ZPh* 4. 241–252, 328–364 (1950).
G. A. Miller, "The perception of speech," *FRJ* 353–359.
—, *Language and communication* (New York, 1951).
—, G. A. Heise, and W. Lichten, "The intelligibility of speech as a function of the context of the test materials," *JEP* 41.329–335 (1951).
R. L. Miller, "Auditory tests with synthetic vowels," *JASA* 25.114–121 (1953).

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

G. Panconcelli-Calzia, *Die experimentelle Phonetik in ihrer Anwendung auf die Sprachwissenschaft* (Berlin, 1924).

G. E. Peterson, "Design of visible speech devices," *JASA* 26.406–413 (1954).

—, "The discrete and the continuous in the symbolization of language," *SJWh* 209–218.

—, "Fundamental problems in speech analysis and synthesis," *PICL* VIII (Oslo, 1957).

—, "The information bearing elements of speech," *JASA* 24.629–637 (1952).

—, "The phonetic value of vowels," *Language* 27.541–553 (1951).

—, "Phonetics, phonemics, and pronunciation: spectrographic analysis," *Georgetown University Monograph Series on Languages and Linguistics*, No. 6, ed. E. Pulgram (1954), 7–18.

—, "The spectrographic analysis of speech," *Georgetown University Monograph Series on Languages and Linguistics*, No. 7, ed. H. Mueller (1954), 62–83.

K. L. Pike, *Phonemics* (Ann Arbor, 1947).

—, *Phonetics* (Ann Arbor, 1943).

R. K. Potter, A. G. Kopp, and H. C. Green, *Visible speech* (New York, 1947).

—, and J. C. Steinberg, "Toward the specification of speech," *JASA* 22. 807–820 (1950).

E. Pulgram, "Phoneme and grapheme: a parallel," *Word* 7.15–20 (1951).

G. Rosen, K. N. Stevens, and J. M. Heinz, "Dynamic analog of the vocal tract," *JASA* 28.767 (A) (1956).

L. O. Schott, "A playback for visible speech," *BTS Lab. Rec.* 26.333–339 (1948).

G. R. Springer, "Language and music: parallels and divergencies," *FRJ* 504–513.

J. C. Steinberg, "Application of sound measuring instruments to the study of phonetic problems," *JASA* 6.16–24 (1934).

R. H. Stetson, *Motor phonetics* (Amsterdam, 1951).

K. N. Stevens, and A. S. House, "Studies of formant transitions using a vocal tract analog," *JASA* 28.578–585 (1956).

—, and —, "Development of quantitative description of vowel articulation," *JASA* 27.484–493 (1955).

—, S. Kasowski, and C. G. M. Fant, "An electrical analog of the vocal tract," *JASA* 25.734–742 (1953).

S. S. Stevens, *Handbook of experimental psychology* (New York, 1951).

—, and H. Davis, *Hearing* (New York, 1947).

C. Stumpf, *Die Sprachlaute* (Berlin, 1926).

F. Trendelenburg, *Einführung in die Akustik* (Berlin, 1950).

E. S. Weibel, "Vowel synthesis by means of resonant circuits," *JASA* 22. 858–865 (1955).

W. S.-Y. Wang and G. E. Peterson, "Segment inventory for speech analysis," *JASA* 30.742–746 (1958).

J. Wiren, and H. L. Stubbe, "Electronic binary selection system for phoneme classification," *JASA* 28.1082–1091 (1956).

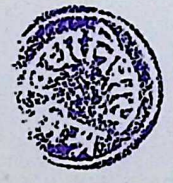
Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Abbreviations

BTS	Bell Telephone System
FRJ	M. Halle, H. G. Lunt, H. McLean, and C. H. van Schooneveld, ed., *For Roman Jakobson* (The Hague, 1956)
IVA	Ingeniors vetenskaps akademien och dess laboratorier
JASA	Journal of the Acoustic Society of America
JEP	Journal of experimental Psychology
MPh	Maître phonétique
MiPh	Miscellanea phonetica
PICL	Proceedings, International Congress of Linguists
PMLA	Publications, Modern Language Association of America
SJWh	Ernst Pulgram, ed., *Studies . . . Joshua Whatmough* (The Hague, 1957)
ZPh	Zeitschrift für Phonetik

Vidya Niwas Mishra
Sanskrit Department
University of Gorakhpur,
GORAKHPUR.



Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji

CC.0 Digitized by Arya Samaj Foundation, Chennai and eGangotri

Vidya Niwas Mishra Collection, Varanasi. Courtesy. Dayanidhi Mishra ji